



Article scientifique

Article

2020

Published version

Open Access

This is the published version of the publication, made available in accordance with the publisher's policy.

---

## Context-dependent multiplexing by individual VTA dopamine neurons

---

Kremer, Yves; Flakowski, Jérôme; Rohner, Clément Adrien; Luescher, Christian

### How to cite

KREMER, Yves et al. Context-dependent multiplexing by individual VTA dopamine neurons. In: Journal of Neuroscience, 2020, vol. 40, n° 39, p. 7489–7509. doi: 10.1523/JNEUROSCI.0502-20.2020

This publication URL: <https://archive-ouverte.unige.ch/unige:150120>

Publication DOI: [10.1523/JNEUROSCI.0502-20.2020](https://doi.org/10.1523/JNEUROSCI.0502-20.2020)

# Context-Dependent Multiplexing by Individual VTA Dopamine Neurons

Yves Kremer,<sup>1\*</sup> Jérôme Flakowski,<sup>1\*</sup> Clément Rohner,<sup>1</sup> and Christian Lüscher<sup>1,2</sup>

<sup>1</sup>Department of Basic Neurosciences, University of Geneva, Geneva, CH-1211, Switzerland, and <sup>2</sup>Clinic of Neurology, Department of Clinical Neurosciences, University Hospital Geneva, Geneva, CH-1205, Switzerland

Dopamine (DA) neurons of the VTA track cues and rewards to generate a reward prediction error signal during Pavlovian conditioning. Here we explored how these neurons respond to a self-paced, operant task in freely moving mice. The animal could trigger a reward-predicting cue by remaining in a specific location of an operant box for a brief time before moving to a spout for reward collection. VTA DA neurons were identified using DAT-Cre male mice that carried an optrode with minimal impact on the behavioral task. *In vivo* single-unit recordings revealed transient fast spiking responses to the cue and reward in correct trials, while for incorrect ones the activity paused, reflecting positive and negative error signals of a reward prediction. In parallel, a majority of VTA DA neurons simultaneously encoded multiple actions (e.g., movement velocity, acceleration, distance to goal, and licking) in sustained slow firing modulation. Applying a GLM, we show that such multiplexed encoding of rewarding and motor variables by individual DA neurons was only apparent while the mouse was engaged in the task. Downstream targets may exploit such goal-directed multiplexing of VTA DA neurons to adjust actions to optimize the task's outcome.

**Key words:** dopamine; encoding; motor action; reward

## Significance Statement

VTA DA neurons code for multiple functions, including the reward prediction error but also motivation and locomotion. Here we show that about half of the recorded VTA DA neurons perform multiplexing: they exploit the phasic and tonic activity modes to encode, respectively, the cue/reward responses and motor parameters, most prominently when the mouse engages in a self-paced operant task. VTA non-DA neurons, by contrast, encode motor parameters regardless of task engagement.

## Introduction

In the interest of survival, animals navigating through the environment adapt their behavior when reward contingencies change. Predicting a reward based on prior outcome contributes to the necessary learning process through a phasic response (i.e., a burst of high frequency spikes) in midbrain DA neurons. Such reward prediction error (RPE) signaling promotes the formation

of cue-reward associations (Schultz et al., 1997). RPE signaling can be observed in Pavlovian conditioning (Saunders et al., 2018) where the animal has little agency. This is particularly apparent for overtrained animals where the phasic response shifts from the reward to the cue during the formation of a novel cue-reward association. RPE coding scales with reward value (Tobler et al., 2005), reward probability (Fiorillo et al., 2003), and timing between events (Hollerman and Schultz, 1998). If preceded by several stimuli, the signal shifts to the earliest reliable reward-predicting cue (Schultz et al., 1993).

In general, RPE computations may be more complex, possibly through the inclusion of hidden states inferred from previous training or subjective valuation (Sadacca et al., 2016; Starkweather et al., 2017). For example, in self-paced operant behavior, the activity of midbrain DA neurons and the ensuing DA transients in the NAC may correlate with operant responding during drug or food seeking (Phillips et al., 2003; Roitman et al., 2004). Additional studies suggest that DA neurons can also code for motivation to work (Hamid et al., 2016; Mohebi et al., 2019), effort (Walton and Bouret, 2019), action initiation (Syed et al., 2016), distance to reward (Howe et al., 2013), and time between stimuli (Soares et al., 2016). This raises the

Received Feb. 20, 2020; revised July 25, 2020; accepted Aug. 3, 2020.

Author contributions: Y.K., J.F., and C.L. designed research; Y.K. and C.R. performed research; Y.K. and J.F. analyzed data; Y.K., J.F., and C.L. wrote the first draft of the paper; Y.K., J.F., and C.L. edited the paper; C.L. wrote the paper.

\*Y.K. and J.F. contributed equally to this work.

C.L. is a member of the scientific advisory board of the Phenix Foundation, Geneva. The remaining authors declare no competing financial interests.

This work was supported by Swiss National Science Foundation FNS310030B\_170266 and European Research Council UET-MESSI-322541. We thank Anthony Holtmaat, Alexandre Pouget, Rui Costa, Wolfram Schultz, Léon Fodoulan, Michael Loureiro, and Ruud van Zessen for helpful comments on the manuscript; all members of the C.L. laboratory for stimulating discussions; and Sebastien Pellat for support in setting up the apparatus for behavioral measurements.

Correspondence should be addressed to Christian Lüscher at christian.luscher@unige.ch.

<https://doi.org/10.1523/JNEUROSCI.0502-20.2020>

Copyright © 2020 the authors

question of whether a single VTA DA neuron encodes multiple behavioral variables at the same time. It has been proposed that the phasic component of DA neuron firing encodes the external events (e.g., a cue and reward), whereas more sustained changes in firing in the same cell, commonly denominated as the tonic firing, would encode movement vigor or motivation (Niv et al., 2007), although this point is still under debate (Mohebi et al., 2019). The simultaneous encoding of rewarding and motor variables in DA neurons is supported by recent findings in head-fixed mice monitoring neural activity with calcium imaging in a virtual-reality environment (Engelhard et al., 2019). In addition, electrophysiological recordings suggest that DA neurons tend to covary with body movements in partially restrained mice, namely, in a spring-suspended basket (Coddington and Dudman, 2018) or on a running wheel (Dodson et al., 2016). Spiking correlations between reward and locomotion have been reported in a freely foraging task in rats (Puryear et al., 2010), whereas tonic and phasic firing have been used to finely resolve the DA dynamics of motivation in a 2-armed bandit task initiated by a cue signal (Mohebi et al., 2019).

However, to the best of our knowledge, no study has probed the activity of optogenetically identified VTA DA neurons in a self-paced operant task where the animals can freely move. Thus, we performed high-resolution single-unit recordings to access encoding of rewarding and motor variables over time during a task elaborated by combining an operant behavior test originally developed to study hippocampal place cell firing patterns (Blanquat et al., 2013; Hok et al., 2013) with food approach (London et al., 2018). In brief, mice had to find an unmarked “trigger zone” (TZ) in the operant box and remain there for 2 s, which would activate a light cue. Once the cue was presented, the animal had 4 s to collect the reward by licking for a drop of fat solution at a spout located at the other side of the box. The animal could engage in the next trial at its own pace. We performed single-unit recordings in the VTA and video recorded the movement of the mouse. We found that individual VTA DA neurons simultaneously encode cue/reward events in the phasic response and motor output in the tonic activity.

## Material and Methods

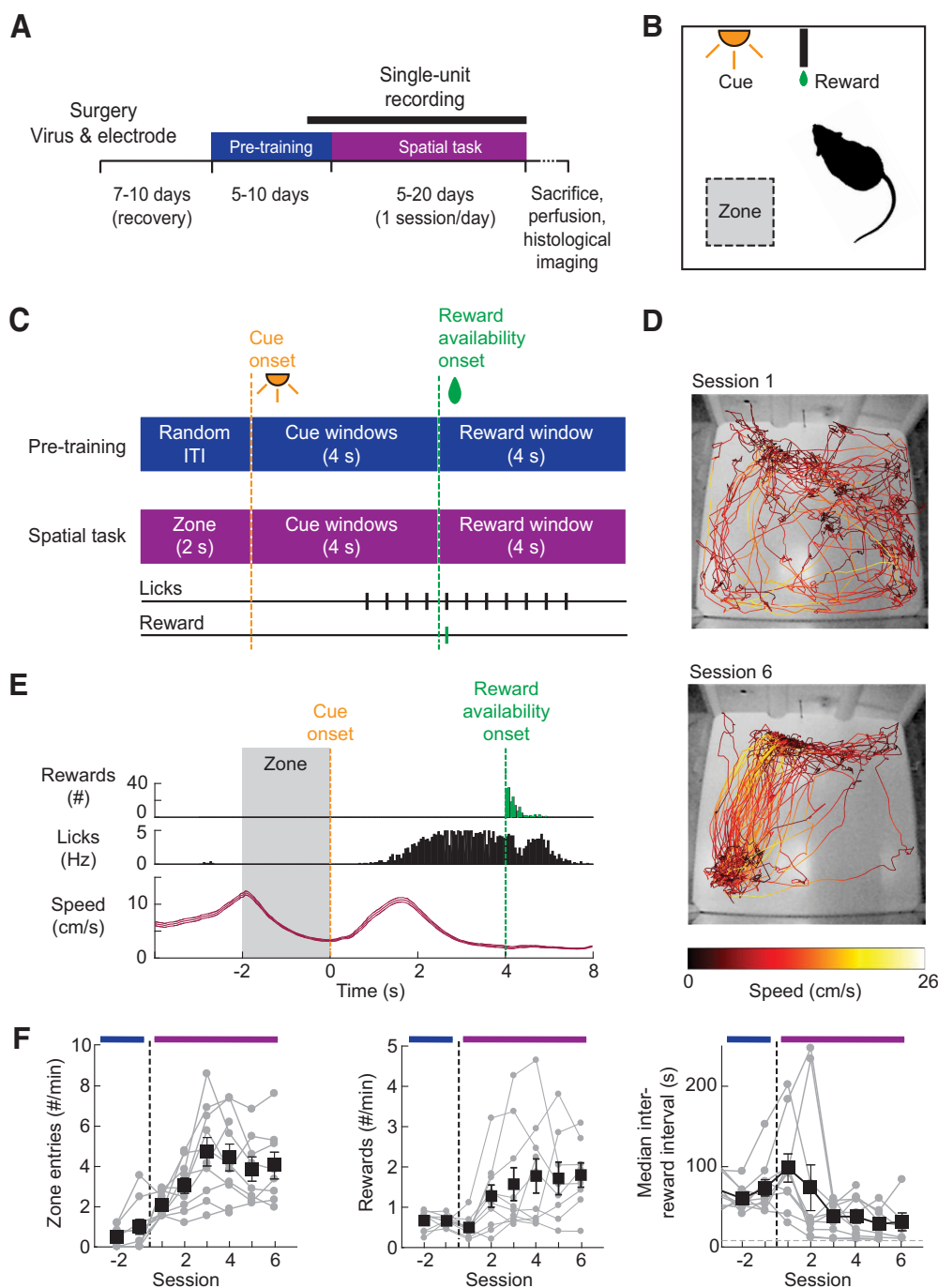
**Animals.** All experiments were reviewed by the institutional ethics committee and approved by the relevant authorities of the Canton of Geneva. Animals were housed individually after implantation to avoid damage to the optrode. During training, they were food-restricted to a single pellet of chow per day (~2 g). The animal's weight did not drop below 85%. All animals ( $n = 10$ ) used in this study were adult Dopamine transporter (DAT)::IRES-Cre knock-in (identifier: Tg(Slc6a3-icre)1Fto, RRID:MGI\_3770172). In DAT-Cre mice, the Cre-recombinase is expressed only in dopaminergic neurons, characterized by the DAT. Only males were used for the electrophysiological recordings to minimize the impact of the weight of the optrode implant.

**Surgical procedures and viral injections.** For recordings, adult male DAT-Cre mice ( $n = 10$ ) were implanted with a custom-built 16-channel optrode mounted into a microdrive (Anikeeva et al., 2011). Sixteen NiCr microwires (tetrode wire 12.5  $\mu\text{m}$ ; Sandvik) were assembled into an electrode bundle, glued to a 250  $\mu\text{m}$  large NA (0.66) optic fiber (Prizmatix, London; [www.prizmatix.com](http://www.prizmatix.com)), and mounted into the microdrive assembly. The tips were electroplated in gold solution ([www.neuralynx.com](http://www.neuralynx.com)) containing multiwall carbon nanotubes (1 mg/ml) ([www.cheaptubes.com](http://www.cheaptubes.com)) (Ferguson et al., 2009) to reach a resistance of 100–150 k $\Omega$  using the NanoZ impedance tester. The completed implant did not

exceed 2 g in weight. During the surgery, mice were anesthetized using isoflurane (~2%), the scalp was injected with lidocaine, and the skull exposed. Three skull screws were inserted and a craniotomy performed above the VTA. The dura was carefully removed, and 400–600 nl of cre-dependent channelrhodopsin (ChR2) (AAV5-EF1a-DIO-hChR2(H134R)-eYFP-WPRE-pA; UNC Vector Core) was slowly injected into the VTA (AP:  $-3.2 \pm 0.2$ ; ML:  $0.8 \pm 0.2$ ; DV:  $-4.2$ ). Then, the optrode was lowered slowly into the brain using a micromanipulator and implanted just above the VTA (AP:  $-3.2 \pm 0.2$ ; ML:  $0.8 \pm 0.2$ ; DV:  $-3.8$  to  $-4.0$ ). The space between the brain surface and the implant was first sealed using silicone elastomer (Kwik-Cast, World Precision; [www.wpi-europe.com](http://www.wpi-europe.com)), then secured using super glue and dental cement. The ground electrode was inserted into the posterior part of the contralateral hemisphere devoid of brain vessels.

**Behavioral training.** After the surgery, animals were given at least 1 week to recover (experimental timeline in Fig. 1A). For recording experiments, DAT-Cre mice were progressively food-restricted down to 1 pellet of chow per day to reach ~85%–90% of their initial weight. During one habituation session in the behavioral apparatus (Med Associates), they located the drinking spout and had free access to the fat solution (solution of lipofundin 5% in water). Then mice started pretraining to associate a light cue with the availability of a liquid reward. If during the reward window mice licked against a custom-made piezo-based lickometer, they obtained a single drop of liquid reward. The analog piezo lickometer signal was sent to an Arduino board, which applied a threshold and simultaneously sent a digital signal to the MedPC software ([www.med-associates.com](http://www.med-associates.com)) and to our electrophysiology recording Plexon system for high temporal resolution time-stamping of lick detection (OmniPlex Neural Data Acquisition System; [www.plexon.com](http://www.plexon.com)). Lick bursts (LBs) were defined as starting when the interlick interval (ILI) dropped below 0.5 s and ended once the ILI exceeded 2 s. The duration of the cue was 4 s and the reward window lasted 4 s. Both remained unchanged across all training stages. During pretraining, the average random intertrial interval between two consecutive cues was increased across pretraining sessions from an average of 45–65 s during the last session (Fig. 1C). Animals underwent pretraining until they exhibited good success rates ( $>0.8$ ) of reward per cue while simultaneously reducing their licking between subsequent cues. During both the pretraining and the operant task, AnyMaze (Stoelting; [www.anymaze.co.uk](http://www.anymaze.co.uk)) or CinePlex (Plexon), were used to online video track the animal's position (defined by its center of gravity). A fixed  $4 \times 4 \text{ cm}^2$  zone, identical for all mice, was defined inside the  $22 \times 22 \text{ cm}^2$  operant chamber (Fig. 1B). The zone limits are not visible for the mice. Digital signals were sent to MedPC (MedAssociates) to trigger the light cue and reward delivery, if the mouse spent 2 s inside the TZ as required by the task. Additionally, all zone entries, licks, cues, and rewards were recorded with high temporal resolution by our electrophysiology recording system. We did not observe distinct sign or goal-tracking behavior during the cue-reward time interval. The animal's performance was quantified using three different metrics during the task (Fig. 1F). On top of the usual reward rate, the number of zone entries and the median inter-reward interval were used. The median inter-reward interval (Fig. 2) is more accurate than the reward rate since it is not sensitive to outliers (i.e., small number of reward collection events being far apart from each other) crucial for long behavioral session where the animal is free to make breaks in the task execution. The shorter the interval, the better the animal's performance.

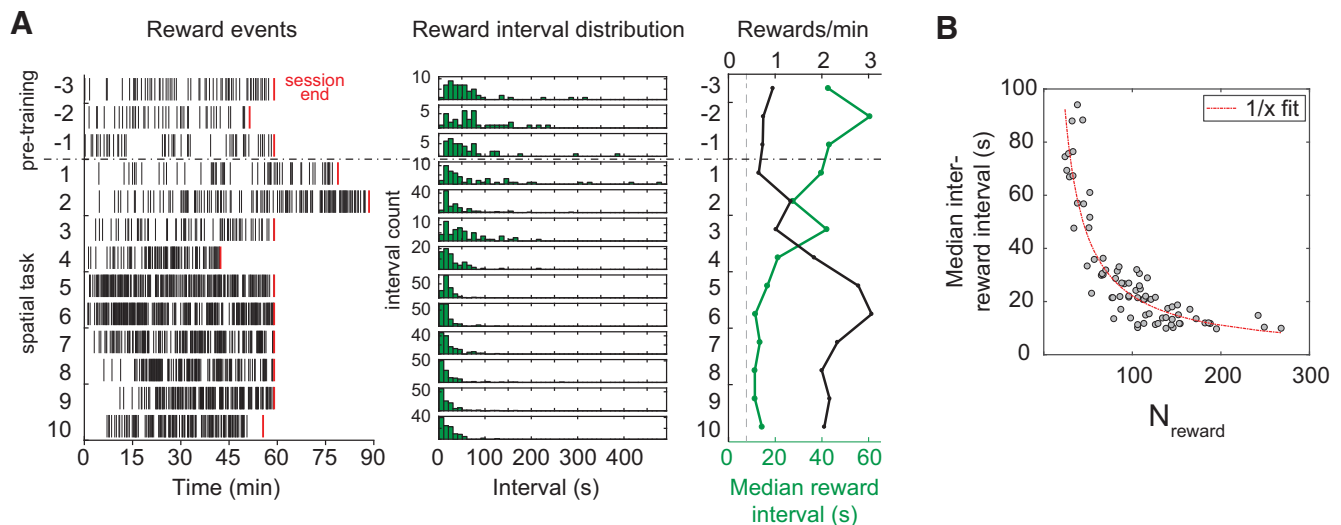
**Electrophysiological recordings and spike sorting.** We started recordings during the last pretraining sessions when mice



**Figure 1.** Performance in self-initiated, operant goal-directed task. **A**, Experimental timeline. **B**, Schematics of the task. The  $4 \times 4$  cm<sup>2</sup> fixed trigger zone (TZ), in a  $22 \times 22$  cm<sup>2</sup> box, is depicted here in gray but invisible for the mice that have to make the association between their spatial position and the cue to get a single drop of fat solution at the spout. **C**, During pretraining, mice have to associate a randomly occurring 4 s light cue to the availability of the liquid reward. Once this association had been learned, we switch to the operant task during which mice have to wait for 2 s in the zone to trigger the light cue indicating that the reward can be collected. **D**, Track plots from 10,000 video frames ( $\sim 6$ –7 min) showing the locomotor pattern of the same mouse at the first and sixth sessions of the spatial task. **E**, Behavioral variables aligned to the cue onset recorded during Session 6: rewards collected during the 4 s reward window (green, 100 ms bin), individual licks (black, 100 ms bin), and speed (dark red, mean  $\pm$  SEM) are aligned to the cue. **F**, Behavioral performance of all mice ( $n = 10$ ) for the last pretraining (–2 to –1) and the operant task (1–6) sessions. Black squares and error bars represent mean  $\pm$  SEM across mice.

reached the criterion to move to the operant task (Fig. 1A). If a neuron exhibited DA-like firing patterns or responded to the opto-identification protocol, the mouse was switched to the operant task on the next day. Neuronal activity was recorded at 40 MHz sampling rate and all behavioral variables timestamped by our recording system (Plexon). To allow for unrestrained locomotion and to reduce weight, we used a dual LED and 16-channel commutator (Plexon). The distance between the

headstage and the commutator was optimized for each animal. Recording sessions lasted 60–90 min; simple online sorting was performed to monitor recording quality. For analysis, offline spike sorting was performed using WaveClus (Quiroga et al., 2004) (University of Leicester, MATLAB code; [https://github.com/csn-le/wave\\_clus](https://github.com/csn-le/wave_clus)). A frequency of 150 Hz was used to separate the wideband signal into local field potentials and spike bands. Positive or negative thresholds were applied depending



**Figure 2.** Median inter-reward interval as performance metric. **A**, Performance of one mouse across learning stages. Left, Temporal distribution of rewards for the last three pretraining sessions (Sessions  $-3$  to  $-1$ , top) and across 10 sessions of the operant task (Session 1–10). Red ticks represent the end of each session. Middle, Distribution of inter-reward time intervals across sessions (green, 10s bin), showing a gradual narrowing around short durations. Right, Median inter-reward interval (green) and average reward rate (black) across sessions. **B**, Median reward interval plotted against the number of rewards obtained during the same session (gray dots) shows that a small median interval does reflect better performance and is not because of a low number of rewards. Red dashed line indicates a  $1/x$  fit. The median inter-reward interval is insensitive to slow initiations or occasional breaks, which is why we used it to assess the performance of the animal.

on the observed waveforms. Spike sorting results were visually inspected, the optimal temperature parameter in WaveClus selected, and minimal manual correction applied. The L ratio, a measure to assess the quality of isolation of each unit, was computed using MClust (University of Minnesota, MATLAB code: <http://redishlab.neuroscience.umn.edu/MClust/MClust.html>). Only units with an L ratio  $< 0.05$  were further analyzed (Schmitzer-Torbert et al., 2005; Hill et al., 2011). After each session, the microdrive was used to lower the optrode by  $\sim 50$   $\mu\text{m}$  and the same procedure applied on the next day.

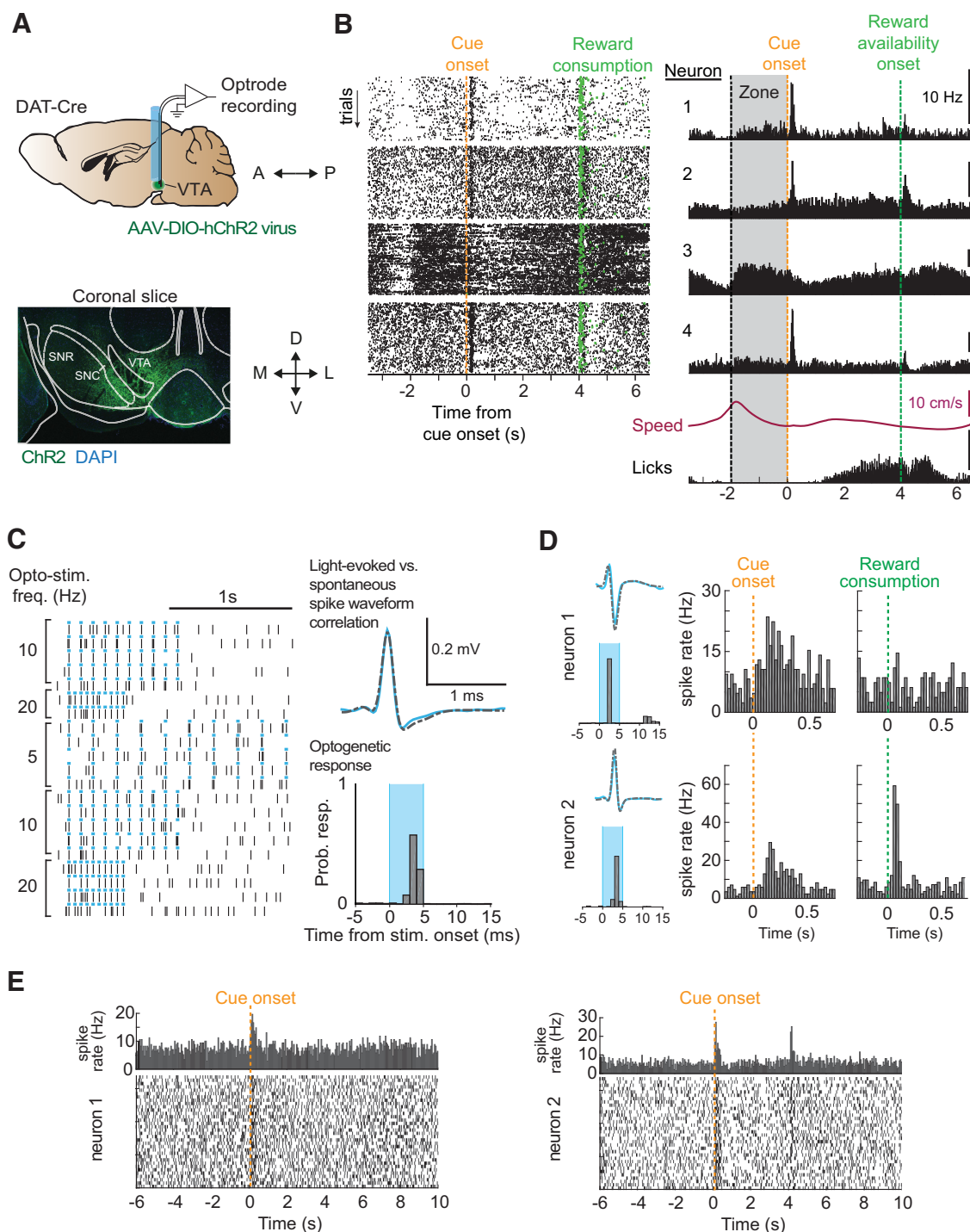
**Optogenetic identification of single units.** After each recording session and before the optrode was lowered for the next day, we delivered blocks of 10 pulses of 5 ms using a LED (Plexbright LED module blue 465 nm, Plexon driven by a Master-8 stimulator, A.M.P.I.) at frequencies of 1, 2, 5, 10, and 20 Hz (Fig. 3C,D). Light intensities ranged from 1 to 5 mW. Light-evoked spikes were detected during a 6 ms window after light onset. A single unit was validated as an optogenetically identified DA neuron if its response rate was  $\geq 0.8$  and the correlation between the average light-evoked waveform and the average of all other spike waveforms produced during the entire recording session was  $\geq 0.85$ .

**Postmortem histology and imaging.** After the optrode had crossed the VTA, mice were killed with a lethal injection of pentobarbital (150 mg/kg) and perfused with 4% of PFA in cold PBS. After fixation, the optrode was slowly retracted using the microdrive and the brain extracted. Coronal brain slices of the VTA were cut 80–120  $\mu\text{m}$  thick (Fig. 3A). To quantify the specificity of ChR2 expression, slices were immunostained for TH. Briefly, slices were permeabilized with 0.1% Triton in TBS-Tween 20 (TBST), blocked using 1% BSA in TBST, and incubated with primary anti-TH antibody (dilution 1:500; AB152, Merck Millipore). At 24 h later, slices were rinsed and incubated with secondary donkey anti-rabbit IgG antibody (dilution 1:500; AP182C, Merck Millipore) diluted in 1% BSA TBST for 1–2 h. After each step, samples were washed with TBST. Slices were mounted and covered on microscope slides using Fluoroshield

mounting medium with DAPI (Abcam, ab104139), low-resolution imaging was performed using either a  $5\times/\text{NA } 0.25$  or  $10\times/\text{NA } 0.45$  objective in a slide scanner (Mirax or AxioScan Z1; Carl Zeiss). High-resolution stacks were acquired in a confocal fluorescence microscope (LSM800 using  $40\times/\text{NA } 1.4$ , Carl Zeiss or A1r Spectral  $40\times/\text{NA } 1.3$ , Nikon) (example in Fig. 4A). In recorded animals, the electrode track could easily be located in 1–3 brain slices and was mapped to the stereotaxic coordinates of the slice where the track was most prominent (Fig. 4B). For cell counting, confocal high-resolution imaging stacks ( $z$  steps of 2.5  $\mu\text{m}$ ) were acquired; then cells expressing ChR2 and those stained for TH were counted manually using ImageJ (National Institutes of Health; <https://imagej.nih.gov/ij/>).

**Event-related neuronal responses for the clustering.** For each neuron, spikes were binned into a regular time grid of 250 ms time bins. Neurons with a mean firing rate  $< 1$  Hz as well as maximal firing rate  $< 3$  Hz were excluded, leaving us with 215 single units ( $n = 10$  mice). For each neuron, we constructed five event-related profiles around cue onset, reward consumption, the mouse's lick, speed peak, and acceleration peak. The increase or decrease of each time bin from the baseline (defined as  $[-3.5; -3.25]$  for the cue onset,  $[2.5; 3.0]$  for reward consumption,  $[-1.0; -0.75]$  for lick onset,  $[-2.0; -1.75]$  for the speed and the acceleration peaks) was quantified using an area under a receiver-operating characteristic curve (auROC) method (Cohen et al., 2012). In brief, auROC values are between 0 and 1; values  $< 0.5$  indicate a decrease relative to the baseline, whereas values  $> 0.5$  reflect increases in firing relative to the baseline (blue/black/yellow heatmaps in Figs. 5 and 6A). To visualize the absolute neuronal firing rate (Fig. 6B,C), we computed for each neuron the corresponding continuous spike density function (cSDF) by convolving the spikes with an EPSP kernel (Wallisch et al., 2014) using the algorithm proposed by the Schall Lab (Hanes et al., 1995) (Vanderbilt University, MATLAB code; [www.psy.vanderbilt.edu/faculty/schall/scientific-tools/](http://www.psy.vanderbilt.edu/faculty/schall/scientific-tools/)).

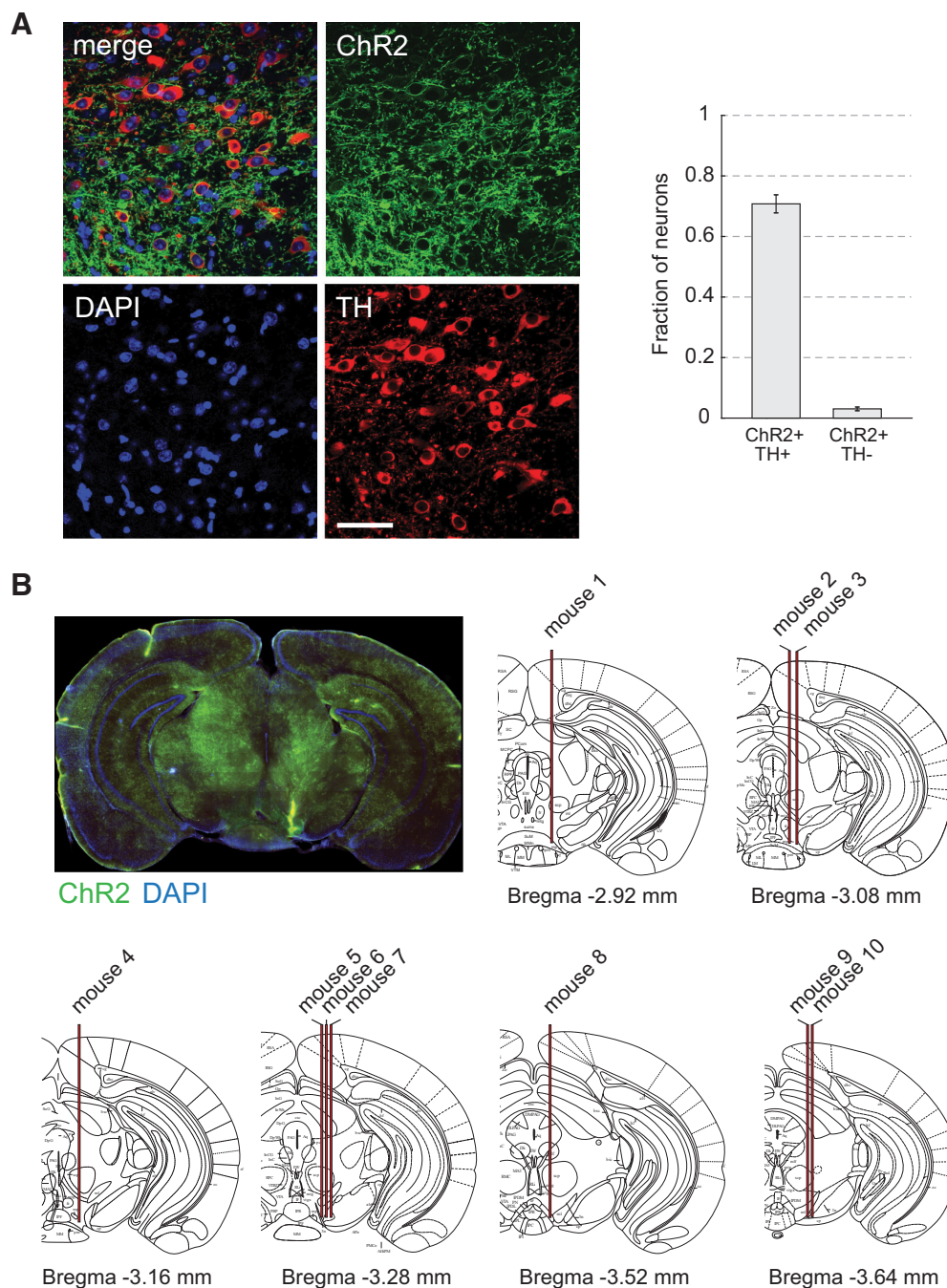
**Clustering of neuronal firing patterns.** We applied a two-step clustering strategy to the 215 neurons ( $n = 10$  mice) to isolate the



**Figure 3.** VTA *in vivo* recording and optogenetic tagging. **A**, Top, Schematics of the single-unit recordings. Custom-made optrodes were implanted into the left VTA of adult DAT-Cre mice ( $n = 10$ ) that were injected with virus expressing floxed ChR2. Bottom, Example of a coronal slice showing ChR2 expression (green) and DAPI staining (blue) in the VTA. **B**, Example of four simultaneously recorded units. Left, Raster plots for each individual neuron aligned to the cue. Right, Corresponding average spiking histogram showing the diversity of firing patterns during the same session along with the speed of the mouse and licking (below). Neurons 1, 2, and 4 were subsequently classified as DA neurons. **C**, Optogenetic identification of DA neurons. Left, 5 ms blue light pulses at 5, 10, and 20 Hz. Neurons exhibiting a response rate  $\geq 0.8$  spikes/pulse during the 6 ms window after light onset and showing a waveform correlation  $\geq 0.85$  were identified as light-responsive (bin size opto = 1 ms). **D**, PSTHs of two DA neurons (bin size: 25 ms), which are light-responsive (neuron 1: opt.resp. = 0.819, waveform corr. = 0.952; neuron 2: opt.resp. = 0.8, waveform corr. = 0.996). They were recorded simultaneously during the same operant task session but display different responses to external events. Both are aligned to the cue onset (left PSTHs) and the reward consumption (right PSTHs). **E**, Raster plots and PSTHs of the two previously optogenetically identified DA neurons recorded during the same session of the operant task (both aligned to the cue). While the first neuron responded exclusively to the cue, the second neuron showed also a response to the reward (as seen in **D**, the response to the reward is stronger).

ones within the VTA. Before the clustering, we applied a dimensional reduction to the auROC profiles (i.e., displayed in the heatmap of Fig. 5 for the cue, reward, lick, speed, and acceleration events) using an independent component analysis (fastICA

MATLAB package, Aalto University; <http://research.ics.aalto.fi/ica/fastica/>). We then applied a hierarchical clustering, using the standardized Euclidean distance metric and ward linkage method, to 18 combined independent component analysis

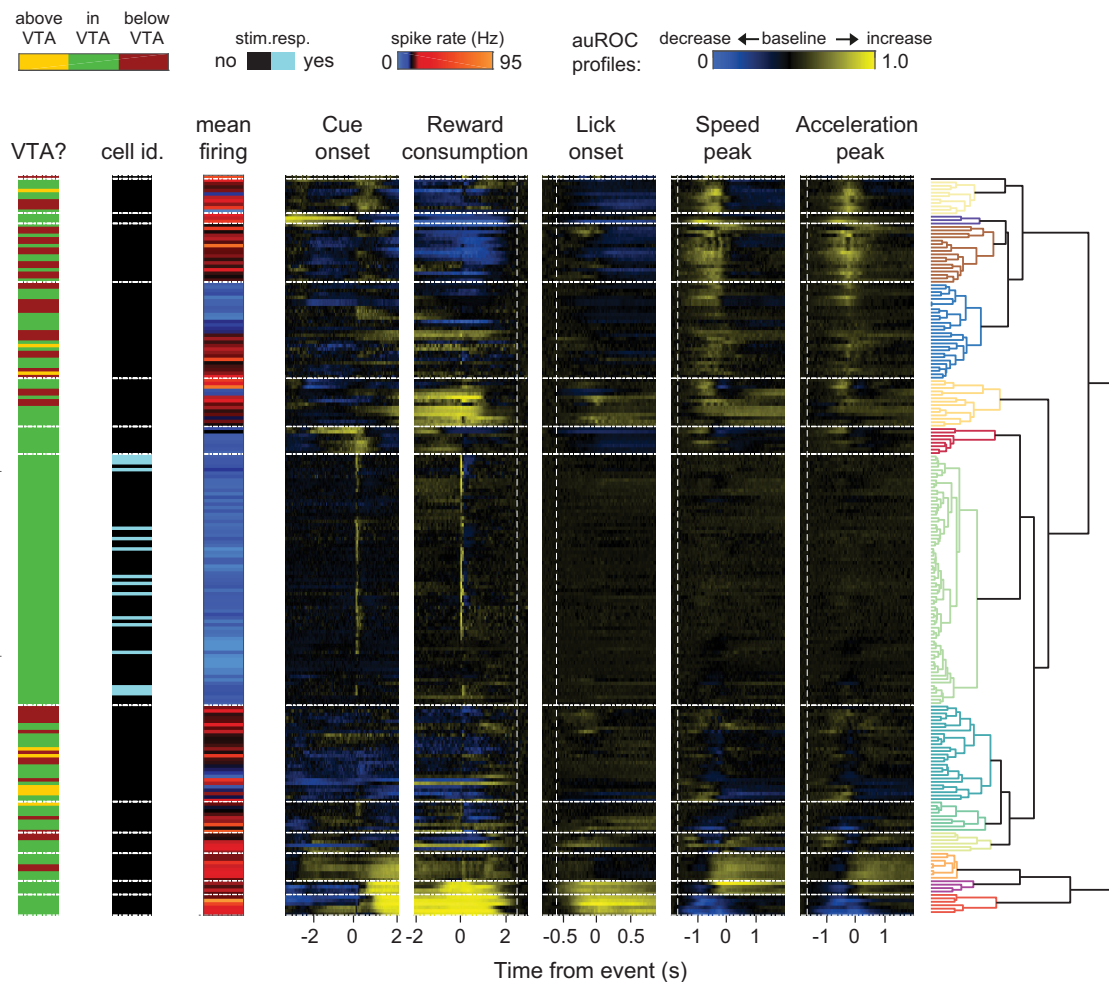


**Figure 4.** Electrode tracks and ChR2 expression in postmortem histology. **A**, Specificity of ChR2 expression compared with TH staining ( $n = 2$  animals). Left, Example section of high-resolution confocal image of combined ChR2 expression (green), TH<sup>+</sup> (red), and DAPI staining (blue). Scale bar, 50  $\mu$ m. We found a high efficiency of transfection of TH<sup>+</sup> neurons by ChR2-eYFP (70.8% of ChR2<sup>+</sup>/TH<sup>+</sup>; 944 of 1328 neurons) and a low proportion of leakage (3.1% ChR2<sup>+</sup>/TH<sup>-</sup>; 43 of 1328 neurons). Bar graph indicates mean  $\pm$  SEM. **B**, Electrode tracks reconstructed after the end of the experiment. Top left, Example of a slice showing ChR2-eYFP expression (green) and DAPI staining (blue). Scale bar, 2 mm. The electrode track can easily be seen in the lateral part of the VTA. Right and bottom, Individual electrode tracks from all animals traced on corresponding brain atlas coronal schema (Paxinos and Franklin, 2001).

components plus the mean firing rate taken across the entire session (i.e., a  $19 \times 215$  matrix). At this stage, the information obtained from the optogenetic identification was not considered. The clustering quality was assessed with the mean silhouette index and cophenetic correlation coefficient. We identified the putative DA cluster, characterized by typical phasic responses to the cue and reward, which identity was confirmed by the fact that all optogenetically identified DA neurons ( $n = 16$ ) fell into it (Cluster 8 in Fig. 5). We exploited the fact that we were moving the optrode across the VTA from session to session with the

microdrive to further refine our set of neurons: we only accepted the ones located between the first and the last DA neuron in a given animal (for 1 animal, we did not find any DA neuron). This resulted in 159 VTA neurons ( $n = 9$  animals), which were subjected to the same clustering procedure yielding this time 9 clusters (Fig. 6A).

**Definition of DA neuron firing modes.** The burst and tonic firing mode was formalized by Grace and Bunney (1984) with the “80/160 ms” rule. A burst starts each time two successive spikes are separated by an interspike interval (ISI)  $< 80$  ms and ends for

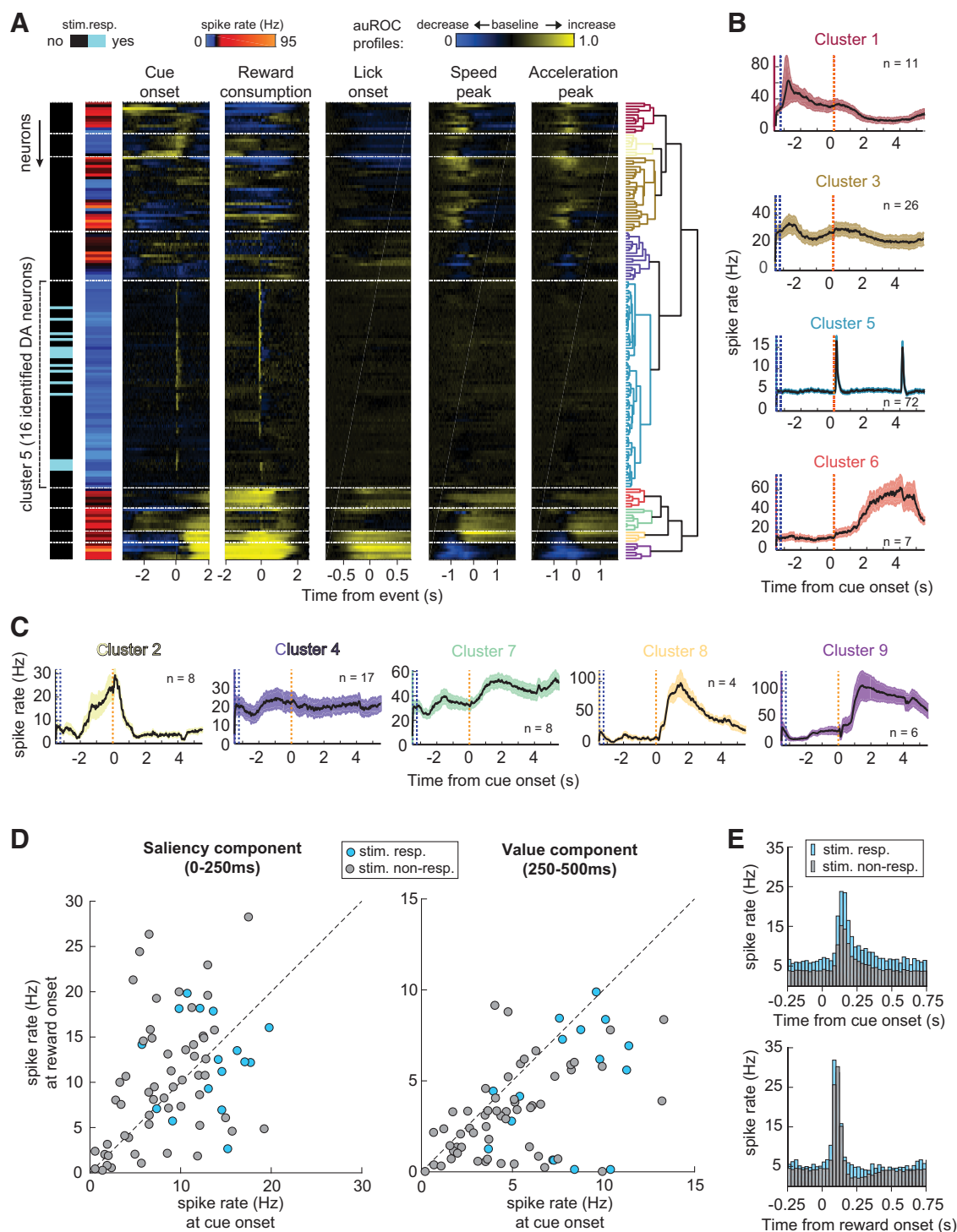


**Figure 5.** Clustering analysis to identify the subset of VTA neurons among all recorded units. The set of 215 recorded units (one per row,  $n = 10$  mice) were aligned to five events of interest (cue onset, reward consumption, lick, speed peak, and acceleration peak) and the relative firing rate changes (250 ms bin) quantified by computing the area under the auROC for each unit following published methods (Cohen et al., 2012). Briefly, the auROC measures to which extend the firing rate at any given time bin is different from the baseline: an excitatory response is characterized by an auROC  $> 0.5$ , whereas an inhibitory response is associated to an auROC  $< 0.5$ . It is well suited for comparing numerically and visually the firing rate of multiple neurons. The auROC profiles, after a dimension reduction, supplemented by the mean firing rate led through a hierarchical clustering to 14 clusters among which the Cluster 8 ( $n = 73$  neurons) contained all optogenetically identified DA neurons (see Materials and Methods). Using this cluster, it was possible to define for each mouse the units recorded in the VTA using the first and last occurrence of a DA neuron (yellow represents above; green represents inside; and red represents below the VTA). The resulting 159 neurons located in the VTA were used for a new hierarchical clustering yielding 9 clusters among which the Cluster 5 contains the DA neurons ( $n = 72$ ; see Fig. 6) considered afterward.

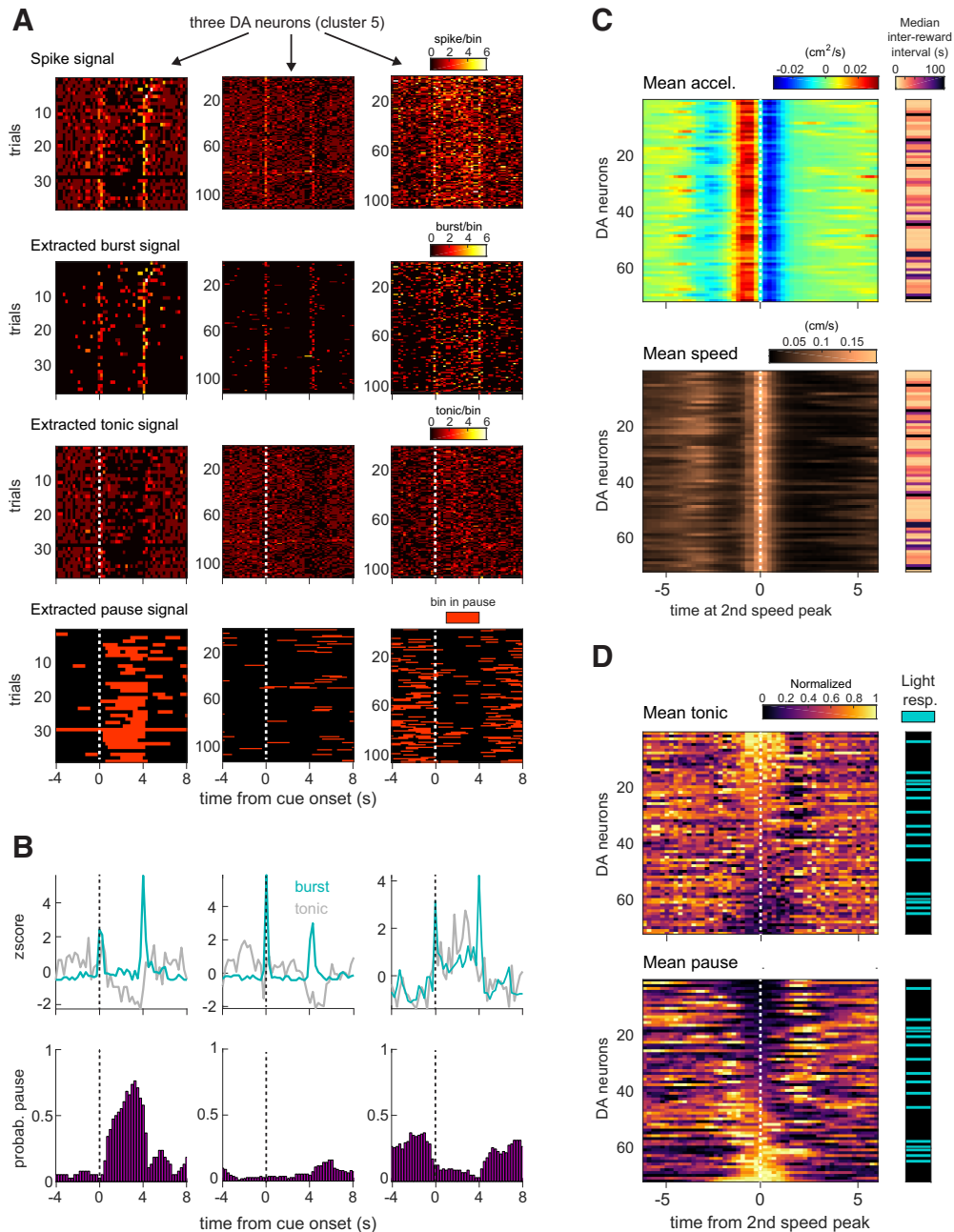
a following spike with an ISI  $> 160$  ms. Since in the freely moving condition, we found elevated firing rates compared with head-fixed animals, we could not directly apply this rule to separate spike bursts from tonic firing. We adapted the 80/160 rule for freely moving animals by first calculating the mean ISI across the entire session. Then, the burst onset was obtained for two spikes separated by an interval shorter than one-third of the mean ISI, and the offset was reached for a following spike with an interval two-thirds of the mean ISI. This resulted in burst detection thresholds similar to previous studies when DA neurons fired at 4–5 Hz, but scaled for neurons with higher firing rates. All spikes not attributed to a spike burst were tagged as tonic spikes. To detect spike pauses, we plotted the histogram of ISIs of all tonic spikes (ISI<sub>tonic</sub>). We then fitted a Gaussian in the interval  $[0; 3]$  of this distribution and computed the time at which the difference between the Gaussian fit and the actual distribution was largest. We used this value as the threshold for pause detection. If this threshold was smaller than  $2 < \text{ISI}_{\text{tonic}} >$ , it was set to  $2 < \text{ISI}_{\text{tonic}} >$  considered as the smallest pause size. In subsequent analyses, time bins were considered to be in a pause if they were between the spike starting a pause and the spike ending a pause. We then

computed the pause probability per bin for correct and error trials. Examples of spike signals split into burst/tonic/pause components are presented in Figure 7A, B.

**Behavioral modules.** To characterize more precisely behaviors during each session, we defined seven modules based on the recorded behavioral variables: resting, move, run, licks, rewarded licks, run to zone, and run to spout. Transition points between resting and moving periods were fixed at 2.1 cm/s. We then computed the mean speed value for move intervals and used a move-run threshold of 2.8 cm/s, thereby defining the resting, moving, and running modules. LBs were detected as described above, starting with an ILI  $\leq 0.5$  s (LB onset) and ending when ILI  $\geq 2$  s (LB offset). We defined rewarded LBs as those whose onsets were occurring between a cue and a reward. All other LBs were considered as unrewarded LBs. Finally, we computed the distance to the reward and to the zone to further specify whether run modules were directed toward the zone or toward the spout. Examples of the modules detected across an entire session are displayed in Figure 8B.



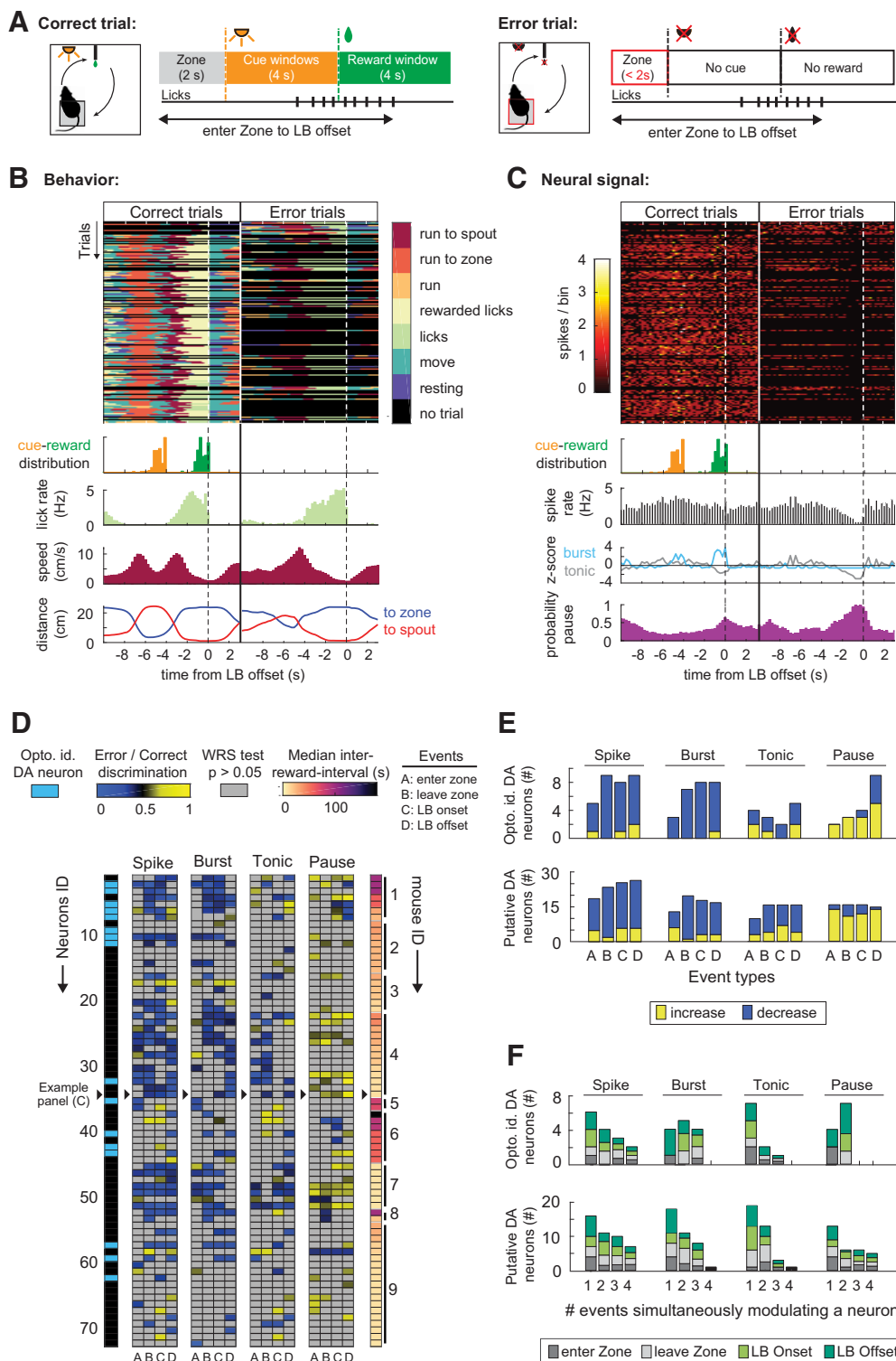
**Figure 6.** Clustering analysis on units in the VTA. **A**, Hierarchical clustering of 159 VTA single units (one per row,  $n = 9$  mice) based on their auROC spiking profiles (yellow represents increase from baseline; blue represents decrease from baseline), for five events and their mean firing rate (for details, see Materials and Methods; Fig. 5). All 16 optogenetically identified neurons fell into Cluster 5, which exhibits large phasic responses around the cue and reward. **B**, **C**, Average continuous spike density function, obtained by convolving the spikes with an EPSP kernel (see Materials and Methods), aligned to the cue (mean  $\pm$  SEM) for functional cluster response patterns. Colors of the SEM correspond to colors of the clusters in the dendrogram in **A**. Some clusters display task-related modulations, for example, speed peaks (1 and 3), cue/reward cluster (5), or licking (6). Some clusters (i.e., 2, 4, 7, 8, and 9) show complex firing modulation, which cannot easily be linked to a single behavior. The auROC profile for the motor events of Cluster 5 (lick, speed, and acceleration peaks) are sampled from the entire session without distinguishing when mice are performing the task (i.e., from the zone entry to the reward collection), leading to weak noisy mean signals. **D**, Average cue versus reward responses during the saliency (0–250 ms) and value (250–500 ms) postevent window for all individual neurons of the Cluster 5 (bin size PSTH = 25 ms). Saliency and value windows are defined according to Schultz's two-component phasic DA response model (Watabe-Uchida et al., 2017). The spike rate between light-responsive and light-nonresponsive neurons is undistinguishable. **E**, Average response histograms of all neurons of Cluster 5, namely, the optogenetically identified (stim. resp.) and putative (stim. non-resp.) DA neurons, to cue onset (top) and reward consumption (bottom) showing average responses to both events. The response to the reward is generally stronger than to the cue.



**Figure 7.** Tonic spike variation during cue-reward interval concomitant with locomotion for DA neurons. **A**, Heatplot of spike, burst, tonic, and pause signal (see Materials and Methods) from three example DA neurons (each column) displaying different responses when aligned to the cue (250 ms bin; baseline [−1, 0]). **B**, The z-scored average burst and tonic firing (blue and gray, top) and mean pause rate or pause probability (purple, bottom) for the same three neurons. During the cue-reward interval, where the animals run from the TZ to the reward spout, the left neurons pauses, the middle one shows little modulation, and the right one exhibits a strong tonic signal. **C**, **D**, Activity of all DA neurons ( $n = 72$ , from the clustering in Fig. 6A) aligned to speed peak during task execution. **C**, Top, Acceleration averaged over all trials associated to the neural recording. Bottom, Speed. Right, Behavioral performance of the animal. There is no bias toward low/high performance (low median inter-reward interval indicates a high performance). **D**, Normalized tonic signal (top) and pauses (bottom) extracted from the spiking of all DA neurons sorted by the smallest to highest mean pause rate. Right, DA neurons that have been optogenetically identified. These graphs provide visual evidence for modulation of the tonic firing by the motor behavior (for quantitative analysis, see Figs. 8 and 9).

*Comparing neural signal underlying correct and error trials.* Correct trials lasted  $\sim 10$  s: the mouse entered the zone, waited there 2 s until the cue turned on, and then ran toward the spout (or valve) where the reward was delivered from 4 to 8 s after the cue. The mouse spent a few seconds near the spout to consume the reward and then returned to the zone (Fig. 1B). To study VTA DA error signals in our setting, we selected trials with a motor output similar to correct trials during which neither the cue nor reward was delivered because the mouse left the zone

too early (Fig. 8A). Under these conditions, the mouse still ran to the spout and licked, although no reward was delivered. Using the behavioral modules defined above, we defined error trials with similar speeds and trajectories to correct trials. In an error trial, we required the time from the module “run to zone” to the module “run to spout” to last at least 3 s to avoid the case when mice just randomly crossed the zone and ran toward the spout. Moreover, to be included in error trials, mice had to lick and test for reward delivery. Similarly, we used these same behavioral



**Figure 8.** VTA DA neurons pause during error trials. **A**, Description of correct and error trials taking both place between the zone entry and LB offset. The animal failed to perform the task when it leaves the zone too early not triggering the cue (right). Only error trials where the animal licked for the reward were considered. **B**, Correct versus error trials are analyzed for one example session. In the neural/behavioral heatmap, the mouse is either in a correct or an error trial at a given line. Colors represent seven behavioral modules (see Materials and Methods). Bottom, Cue (orange) and reward (green) distribution, lick rate (light green), speed (dark red), and the animal's distance to the zone (blue) and to the spout (red) averaged across correct and error trials, respectively. **C**, Neuronal activity heatmap (250 ms bin) of a putative DA neuron split into correct and error trials as defined from the behavioral analysis. On one given line, the displayed spiking corresponds either to a correct or an error trial. Bottom, Average spiking histogram (black), z-scored tonic (gray), and bursting (light blue) activity and probability of being in a spike pause (purple). Plots are aligned to the LB offset. **D**, Correct versus error discrimination of Cluster 5 neurons (optogenetically identified and putative DA neurons,  $n = 72$ ; Fig. 6A). Significantly discriminating neurons are represented by their auROC change between error versus correct trials in spiking, burst firing, tonic firing, and pausing (error/correct discrimination index; see Materials and Methods). Gray represents a nonsignificant error-to-correct change (Wilcoxon rank sum tests,  $\alpha = 0.05$ ). Yellow/blue represents a significantly up/down change (e.g., a strong yellow discrimination index for a pause at the zone exit event indicates that the neuron's firing is pausing during error trials but strongly active during correct trials when aligned to the zone exit). Neurons are grouped per animal and sorted by behavioral performance represented as the median inter-reward interval (Figs. 1F, 2A,B; the shorter the interval, the better the animal's performance). The negative error signal, which manifests as an increase in pause probability for at least one of the four events, is apparent in 63% of optogenetically identified DA

modules to identify correct trials and subsequently validated the procedure using the cue and a reward events, showing that this procedure was capable of extracting correct and error trials based on similar motor output. The ratio of error/correct trials gives an estimate of the operant task success rate of the animal; for example, in Figure 8B, the animal did 103 correct and 23 error trials, so that the success rate is equal to  $103/126 = 0.82$ , confirming that it had well understood the task. This metric is different from the criterion to move from the pretraining to the operant task (ratio of rewards/cues) as well other metrics presented in Figure 1F. Throughout the paper, we use the median inter-reward interval criteria (Fig. 1F, right) to evaluate the animal's performance (Figs. 7–9).

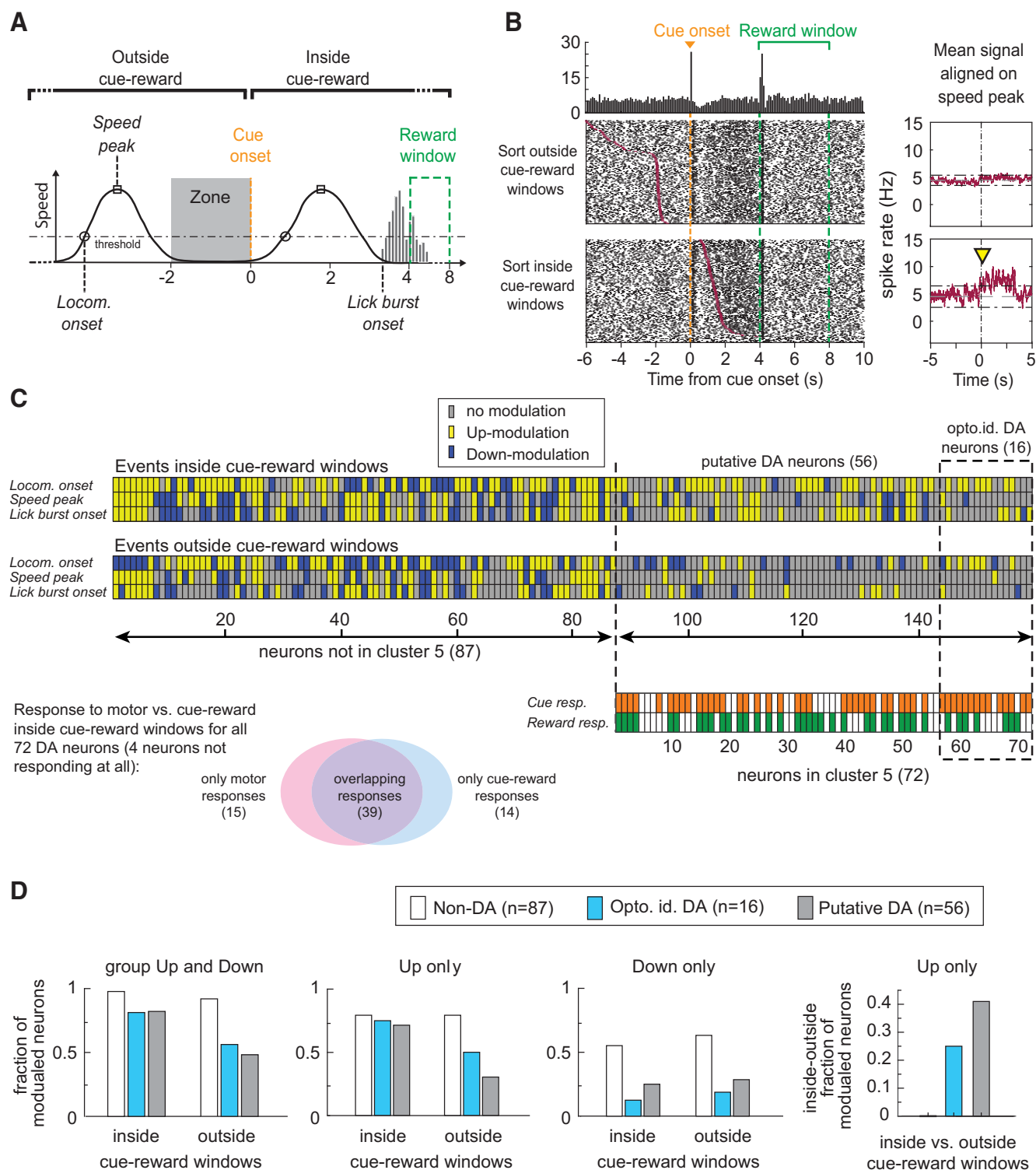
Using the definition of correct and error trials, we then compared their associated neural signal (spike, tonic, burst spiking, and pause) around four events of interest (zone entry, zone exit, LB onset, and LB offset). We applied an auROC analysis (Brodersen et al., 2010) to compare the mean signal in error trials versus correct trials for the four events. The response windows for the auROC comparisons of the different events were as follows: zone entry  $[-0.75; 1.5]$ , zone exit  $[-0.5; 1]$ , LB onset  $[0; 1]$ , and LB offset  $[-0.75; 0.25]$ . To confirm that the relative change was significant, we performed an additional Wilcoxon two-sided rank sum test ( $\alpha = 0.05$ ). For Cluster 5 neurons (Fig. 6A; identified and putative DA neurons have an average firing rate of 4–14 Hz), the analyses were conducted on all spikes, tonic firing, burst spiking, and pauses (Fig. 8D).

**Action related modulation of neuronal spiking inside and outside of cue-reward interval.** To determine which neurons were significantly modulated by goal-directed actions (licking and locomotion), we had first to detect neurons with perievent time histograms (PSTH) displaying a significant activity around LB onsets, locomotion onsets, and speed peaks following a method developed previously (Da Silva et al., 2018). In brief, we selected a window of  $[-5; 5]$  s around these time points, and spikes were time-averaged in a sliding window of 100 ms shifted by 1 ms steps. The baseline was defined as the  $[-5; -4]$  time interval. If the neuronal activity during the response window ( $[-0.5; 0.5]$  for locomotion onset,  $[0; 0.5]$  for speed peak and LB onset) exceeded  $\pm 3$  SDs and remained above (or below) for at least 50 ms, the neuron was considered significantly Up (or Down) modulated by the considered action (example of a significantly larger speed peak in Fig. 10B). The change in firing associated to the motor action (e.g., speed peak) is generally well separated from the firing associated to the cue and reward (see e.g., raster plots in Fig. 10B). Nevertheless, for clear separation, we first removed all movement-related time points closer than 750 ms to an external event before analysis. All these calculations were performed for each neuron inside and outside of the cue-reward window, and the results summarized in a significance heatmap (Fig. 10C; Table 1): yellow represents significant Up modulation; blue represents significant Down-modulation; gray represents no significance.

←  
neurons (10 of 16) and in 41% of putative DA neurons (23 of 56). The neural modulation is not biased toward the animal's performance. Black arrows highlight the example neuron of C, E. Number of neurons significantly discriminating correct versus error trials per event either with an increase or a decrease in the corresponding firing modality for optogenetically identified DA neurons (top) and putative DA neurons (bottom). F, Number and distribution of events, which significantly modulate optogenetically identified DA neurons (top) and putative DA neurons (bottom).

**Encoding analysis of the spiking.** To determine which events and behaviors contributed most to the spiking output, we applied a GLM to predict spiking using a Linear-Nonlinear-Poisson method (Paninski et al., 2007) (see e.g., tutorial example codes; [www.github.com/pillowlab/GLMspiketrainingtutorial](http://www.github.com/pillowlab/GLMspiketrainingtutorial)). The model  $Y = f(\beta X) + \epsilon$  relies on an exponential nonlinear link function  $f$  and a Poisson noise  $\epsilon$  commonly used to predict spiking (Pillow et al., 2005, 2008). To improve between session comparisons, this analysis was restricted to a homogeneous subset of 36 sessions (118 of the 159 neurons). The single units were selected based on the session duration ( $65.9 \pm 3.52$  min, mean  $\pm$  SEM); that is, only units recorded during session with at least 30 min of continuous recordings were selected. The encoding procedure is depicted in Figure 11A. Each activity vector  $Y$  is built from the recorded spiking (one per cell, 250 ms bins). The design matrix  $X$  contained seven covariates (i.e., variables): three events (cue, reward, and zone) and four behaviors (licks, speed, acceleration, and distance to goal) resampled using same bin size as for the spiking. The presence inside the TZ was represented in a binary vector as were cue and valve events, indicated by binary events lasting 2 bins. The remaining continuous variables (licks, speed, acceleration, and distance to goal) were normalized and z-scored. We used a sliding window of 10 bins to predict the spiking at a given time bin. We estimated the regressors  $\beta$  for each neuron using the GLM. To avoid overfitting because of the large number of parameters used to predict the spiking, we resorted to a Lasso and elastic-net regularization (glmnet algorithm for MATLAB, Stanford University; [www.web.stanford.edu/~hastie/glmnet\\_matlab/](http://www.web.stanford.edu/~hastie/glmnet_matlab/)). Pearson's correlation coefficient was used to compare the recorded and predicted spiking (e.g., Fig. 11B). Finally, we temporally shuffled the neural signal 1000 $\times$  by a randomized time offset (50–500 s) and performed the analysis again to check that the prediction is better than chance level (threshold of 2.807 SDs to the shuffled distribution). Accordingly, only neurons whose Pearson correlation was higher than the shuffled one were used (108 of the 118 neurons). To compare the encoding inside versus outside of the task execution time intervals, we concatenate separately the recorded  $Y$  and predicted  $\hat{Y}$  spiking of each neuron for inside the task execution on one hand and outside of the task execution on the other (Fig. 11C). We then computed the correlations for all correct trials' time intervals (ranging from the zone entry to the LB offset) versus outside of correct + error trials' time intervals (Fig. 11D,E). We made the same analysis by considering the correlation for inside versus outside of cue-reward intervals as a cross-check and got similar results. This was expected since all cue-reward intervals are enclosed in correct trial intervals. To determine the number of variables contributing to the spiking of a given neuron, the prediction was performed for each of the possible 127 combinations of the seven variables (i.e.,  $7 + 21 + 35 + 35 + 21 + 1$  from 1–7 variables; see Fig. 12A). In the design matrix  $X$ , removed covariates were set as their mean over the entire session and the regressors  $\beta$  constructed from all seven covariates. We selected the combinations displaying the maximum correlation for each number of covariates and counted the number of times each covariate appeared in these best combinations. We then determined the minimum number of variables (or contributors) necessary to cross 90% of the largest value between the previously computed maximum correlations (Fig. 12A,B). To estimate which variables contributed the most to the spiking pattern, we looked at the relative loss in correlations after removing a single variable from the input to the model (Fig. 12C,D) constructed from six covariates, and therefore computed the following:





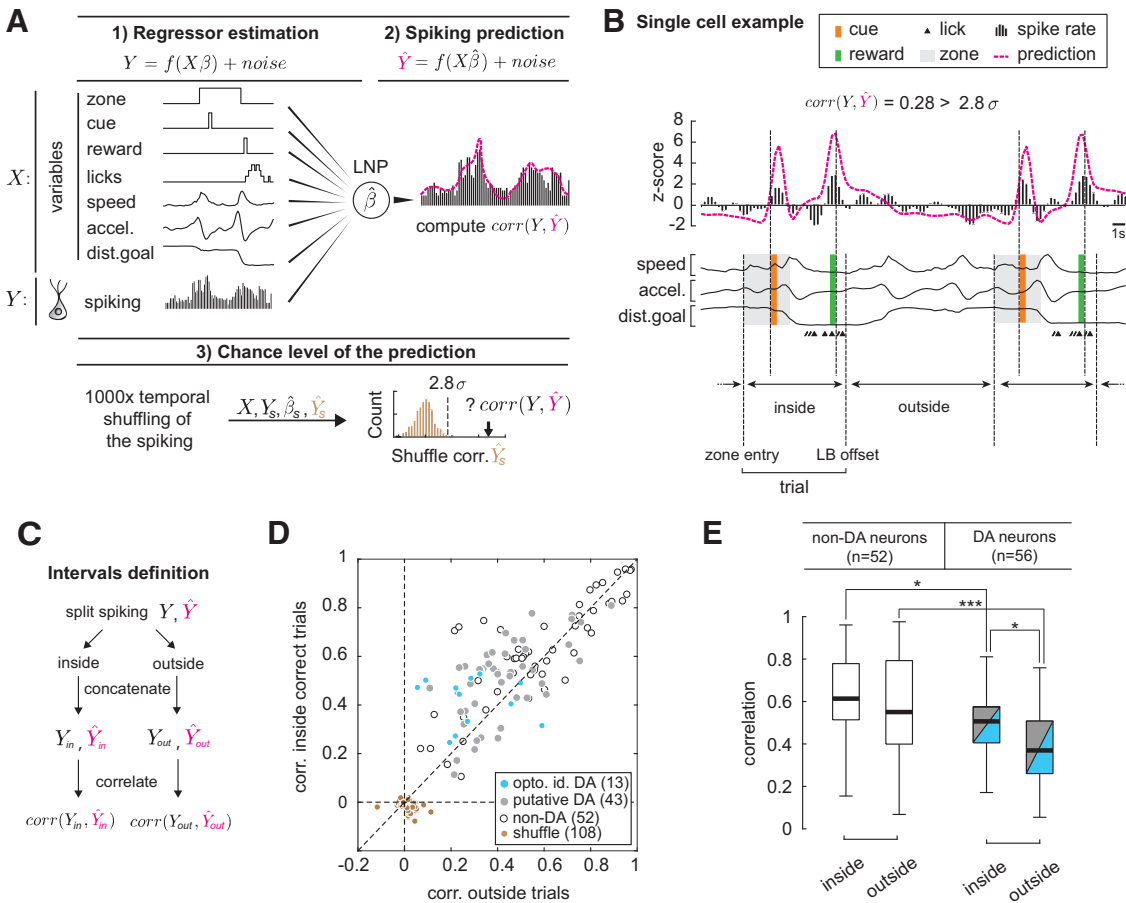
**Figure 10.** Neuronal modulation inside versus outside of the cue-reward interval. **A**, Schematics of three events aligned at the Zone exit used for the comparison below: locomotor onset, speed peak, and LB onset. **B**, Left panels, PSTH (top) and raster plot (bottom) showing the firing of a putative DA neuron aligned to the speed peak (bin size = 100 ms) inside and outside of cue-reward interval. Right panels, Average spike rate aligned at the speed peak outside (top) and inside (bottom) of the cue-reward interval. Spikes of each neuron were first time-averaged in a sliding window of 100 ms shifted by 1 ms steps in a  $[-5; 5]$  s interval around the event. If the modulation during the response window ( $[-0.5; 0.5]$  s for locomotion onset,  $[0; 0.5]$  s for speed peak and LB onset) exceeded  $\pm 3$  SDs and remained above (or below) the baseline modulation (taken from  $[-5; -4]$  s) for at least 50 ms, the neuron was considered significantly Up (or Down) modulated (see Materials and Methods). Bottom, Yellow arrow indicates the start time of significant increase of firing rate. The significant firing change is only observed when the trials are aligned to the speed peak in the cue-reward interval. **C**, Display of significant modulations for all VTA neurons ( $n = 159$ ) for the three variables. Bottom, Identification of phasic activity in all DA neurons and intersection between motor and cue-reward modulation (Venn diagram). Thirty-nine of 72 neurons (54%) show modulation by both modalities, indicating multiplexing in a majority of DA neurons. **D**, The fraction of modulated neurons, depicted individually in **C**, grouped here by cell types and context. The Up modulation is more frequent inside than outside of the cue-reward interval for DA neurons (Cluster 5,  $n = 72$ ; Fig. 6A) compared with non-DA neurons (other clusters,  $n = 87$ ; Fig. 6A) as shown on the rightmost panel. The change in modulation between inside and outside of the cue-reward interval in DA neurons cannot originate from a context difference since non-DA neurons were recorded together with DA neurons but did not show such firing modulations.

Table 1. Summary of the analysis of movement-related modulation of Figure 10

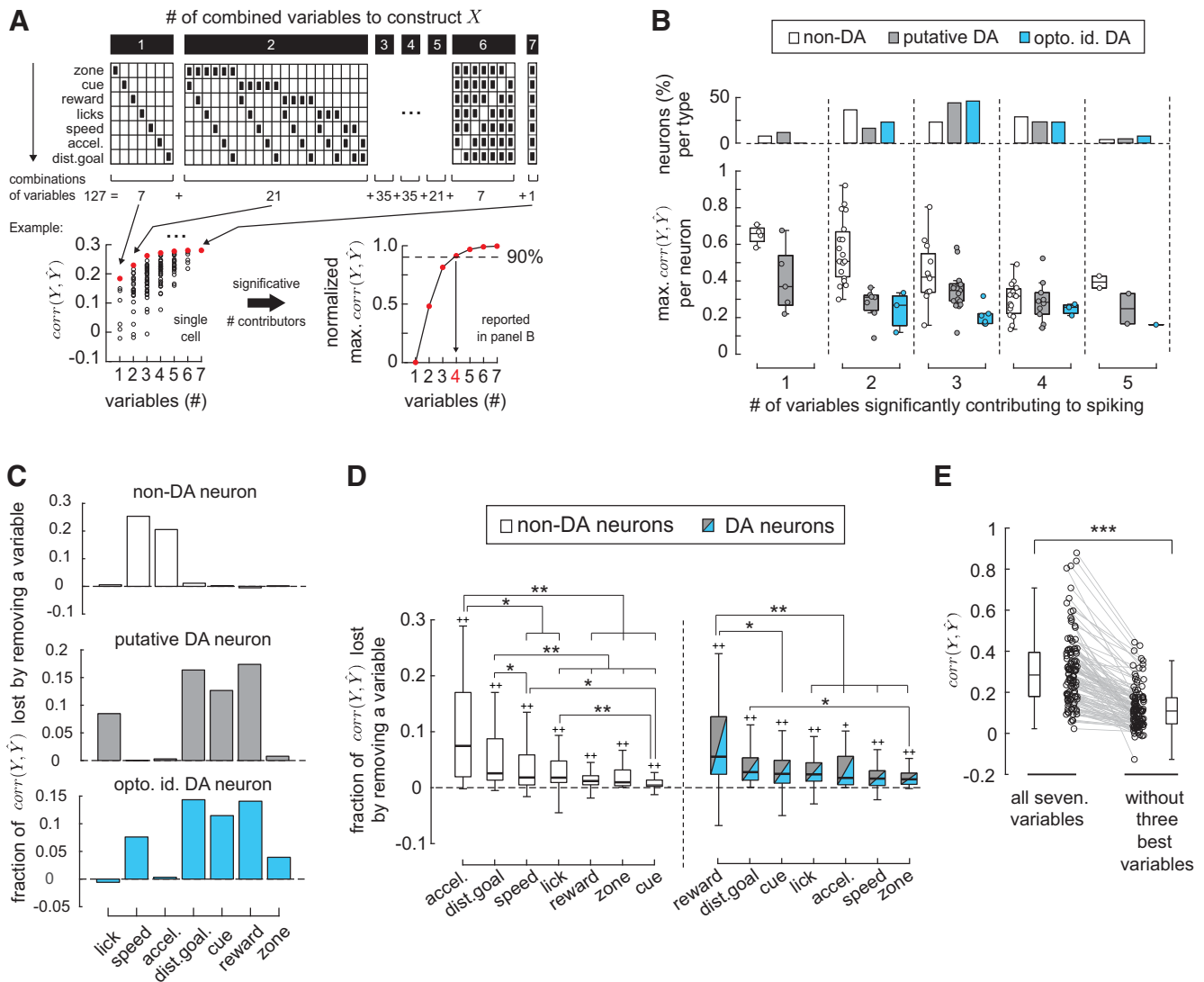
	All events across session						Between cue-reward						Outside cue-reward						Total <i>N</i>	
	Fraction of neurons			Fraction of events			Fraction of neurons			Fraction of events			Fraction of neurons			Fraction of events			Neurons	Events
	All	Up	Down	All	Up	Down	All	Up	Down	All	Up	Down	All	Up	Down	All	Up	Down		
All motor events <sup>a</sup>																				
Non-DA	0.98	0.82	0.74	0.75	0.42	0.33	0.98	0.79	0.55	0.73	0.45	0.28	0.92	0.79	0.63	0.67	0.4	0.28	87	261
Putative DA	0.7	0.55	0.32	0.36	0.24	0.12	0.82	0.71	0.25	0.46	0.36	0.11	0.48	0.3	0.29	0.21	0.11	0.1	56	168
opto. id. DA	0.75	0.63	0.19	0.27	0.21	0.06	0.81	0.75	0.13	0.33	0.29	0.04	0.56	0.5	0.19	0.23	0.17	0.06	16	48
Motor events with onsets further than 750 ms from cue and reward <sup>b</sup>																				
Non-DA	0.95	0.82	0.7	0.72	0.4	0.32	0.87	0.71	0.43	0.64	0.42	0.22	0.93	0.79	0.64	0.66	0.39	0.27	87	261
Putative DA	0.61	0.39	0.39	0.31	0.16	0.15	0.61	0.39	0.29	0.31	0.2	0.11	0.5	0.3	0.3	0.22	0.11	0.11	56	168
opto. id. DA	0.63	0.56	0.19	0.25	0.19	0.06	0.63	0.44	0.19	0.21	0.15	0.06	0.5	0.44	0.19	0.21	0.15	0.06	16	48

<sup>a</sup>Averaged values computed from the individual neuron values as presented in Figure 10C heatmap and used in Figure 10D. The fraction of All neurons indicates how many cells have at least a single significant modulation to motor output. On the other hand, the fraction of Up-modulated neurons shows how many cells display at least one significant increase in spiking, whereas the fraction of Down-modulated neurons shows the number of cells with at least one significant decrease in spiking. The fraction of events counts the number of all significantly modulated, up-modulated or down-modulated events grouped by cell type (non-Cluster 5 or non-DA neurons, putative DA, or opto. id. DA). The total number of events  $n_{\text{events}} = 3 \times n_{\text{neurons}}$  for each cell type.

<sup>b</sup>The same analysis, but excluding all motor events that are closer than 750 ms to the cue or reward (lead to similar results).



**Figure 11.** VTA DA neuronal multiplex encoding is specific to trial execution. **A**, Encoding model. For a given session, the design matrix  $X$  is constructed using seven behavioral covariates (events: zone, cue, and reward; as well as actions: lick, speed, acceleration, and distance to the goal). Each activity vector  $Y$  is built from the recorded spiking (one per cell, 250 ms bins). We first estimated the regressors  $\beta$  for each neuron using a linear-nonlinear-Poisson (LNP) approach; that is, a GLM with an exponential nonlinearity  $f$  and a Poisson noise, for a sliding window of 10 bins with lasso and elastic-net regularization to avoid overfitting. The GLM prediction is compared with the true spiking through Pearson correlations. We proceeded only with neurons whose correlation is larger than chance level for continuous session recording of at least 30 min (see Materials and Methods;  $n = 159 \rightarrow 108$ ). **B**, Top, Encoding example for an optogenetically identified DA neuron. The true (black bars) and predicted (purple dotted line) neural signals have been z-scored and displayed along with all behavioral variables. Bottom, Definition of the time intervals inside a trial (from the zone entry to the LB offset; for details, see Fig. 8A) and outside of a trial used in **C–E**. **C**, We concatenate separately the recorded  $Y$  and predicted  $\hat{Y}$  spiking for inside the task execution on one hand and outside of the task execution on the other. We then computed the associated correlations. **D**, Scatterplot comparing the correlation of inside correct trials versus outside of the trials (excluding the error trials) as defined in **B** and **C**. **E**, Using a Kruskal-Wallis nonparametric unpaired one-way ANOVA with *post hoc* test Tukey-Kramer (regroup putative and opto. id. DA neurons to have groups of comparable size, i.e., number non-Cluster 5 or non-DA neurons = 52 and number DA neurons = 56;  $p < 0.05$ ; \*\*\* $p < 0.001$ ), we show that the variables encoding are significantly stronger during the task execution (zone entry to LB offset) than outside for all DA neurons (putative and optogenetically identified) but not for the non-DA neurons. We also see that the correlations are overall larger for non-DA than DA neurons.



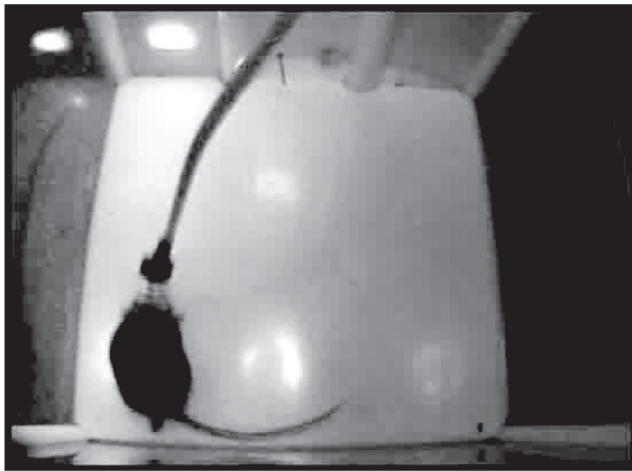
**Figure 12.** Contribution of individual behavioral variables to neural spiking. **A**, Illustration of the method to compute the number of variables significantly contributing to the spiking for each cell (variables combinations and counting). Top, For a given cell, we performed the encoding analysis as depicted in Figure 11A for each combination of variables (127 in total) constructing  $X$  by setting the missing variables to their mean value over time. Bottom, We selected the combination with the minimum number of variables necessary to cross 90% of the largest value among the computed maximum correlations (red dots); in the depicted example, we get four significantly contributing variables (the black arrow, four in red). **B**, Bottom, Maximum correlation for each neuron beyond chance level ( $n = 108$ ) as a function of the number of variables significantly contributing to the spiking following the method depicted in **A**. Top, Percent of neuron per type and per number of significant behavioral variables; considering one or two variables, we have more non-DA (or non-Cluster 5) neurons than identified and putative DA (44% vs 23% and 27%), whereas for three or more variable the tendency is opposite (56% vs 77% and 72%). The average maximum correlation across all neurons decreases with the number of encoded variables (0.53, 0.36, 0.33, 0.27, and 0.26 from 1–5 variables). **C**, The relative importance of a given variable is estimated from the loss in correlation caused by removing it. Histograms represent the relative loss in correlation for each variable (three example neurons, one of each type). Example DA neurons display a similar encoding profile (apart from lick  $\leftrightarrow$  speed). By contrast, the displayed non-DA neuron only encodes two variables, namely, speed and acceleration. **D**, Average loss of correlation for non-DA and DA neurons (combining putative and opto. id. DA neurons to yield groups of comparable size for non-DA neurons,  $n = 52$ , and DA neurons,  $n = 56$ ). The average loss in correlation is significantly different from zero for all variables and neuronal types ( $t$  test with Bonferroni correction;  $^+p < 0.05$ ;  $^{++}p < 0.01$ ) but not for comparisons between variables (for each neuronal type: Kruskal–Wallis nonparametric unpaired one-way ANOVA with *post hoc* test Tukey–Kramer;  $*p < 0.05$ ;  $^{***}p < 0.01$ ). For non-DA neurons, the loss is significantly larger for acceleration versus reward/cue, whereas for DA neurons the loss is significant larger for reward versus acceleration and for cue versus speed. **E**, Comparing the distributions of correlations when the encoder was given all seven covariates to when the three strongest covariates were removed by setting them to a constant value (the average of the covariate). The significant decrease in correlation (two-sample Kolmogorov–Smirnov one-sided test;  $^{***}p < 0.001$ ) confirms that the largest contributing variables are essential for a proper encoding.

box. Importantly, the mouse was free to initiate a new trial by going back to the TZ. The maximum number of rewards was limited only by the duration of the executed sequence ( $\sim 10$  s). Within a few sessions, all mice found the TZ and the reward rate increased, whereas the median inter-reward interval decreased in the first days (Fig. 1F). Since the median inter-reward interval was insensitive to slow initiation or occasional breaks, we chose it as learning metric (Fig. 2A,B). After 3–5 d, the mice executed the task using the shortest trajectories between the TZ and the drinking spout (Fig. 1D,E; Movie 1). Still, even after 5 d, the mice

sometimes left the TZ too early, which led to an error trial. Among the 10 mice, we did not distinguish sign- or goal-tracking behavior. For example, in the speed track plot (Fig. 1D), the fraction of trajectories belonging to speed (orange-yellow lines) are oriented between the light bulb and the spout.

### Identification and classification of VTA neurons

During the operant task, we recorded 215 neurons from 10 mice along the electrode trajectories. The cells exhibited a wide range of firing patterns, responding briefly to the cue and/or the reward



**Movie 1.** Illustration of three successive trials during the operant task: the mice wait 2 s in the zone until the light cue turns on, run toward the reward spout to collect the reward, and then go back to the zone to start the next trial. [View online]

or slowly varying firing pattern (see raster plots and associated PSTHs aligned to the cue onset in Fig. 3B). At the end of each session, before the optrode was lowered for the next day, we optically stimulated (Fig. 3C) and found 16 responding cells across the 10 mice, which are thus confirmed as DA neurons. Two optogenetically tagged DA neurons recorded simultaneously exhibited contrasting responses to the cue and reward (Fig. 3D, E). A classification was then conducted in two steps. First, 215 recorded single units were aligned to five events of interest (cue onset, reward consumption, lick, speed peak, and acceleration peak) and quantified by computing the auROC for each cell following published procedure (Cohen et al., 2012). The auROC indicates how the firing rate can be discriminated from the baseline at any given time bin and is perfectly suited to compare firing patterns of a large number of neurons over multiple events numerically as well as visually (see the heatmaps in Figs. 5, 6A). The auROC profiles subjected to a dimensional reduction combined to the mean firing across the session were processed with a hierarchical clustering (see Materials and Methods) that yielded 14 clusters among which Cluster 8 ( $n = 73$ ; see Fig. 5) contained all optogenetically identified DA neurons. A second classification was then conducted on the cells in the VTA (159 neurons for 9 mice; see Materials and Methods), which were all units recorded between the first and last occurrence of a DA neuron. The resulting hierarchical tree yielded nine clusters among which Cluster 5 contains the DA neurons ( $n = 72$ ; see Fig. 6). In line with previous studies (Cohen et al., 2012; Sadacca et al., 2016; Engelhard et al., 2019), our analysis revealed clusters reflecting motor output, which were modulated by locomotor speed or acceleration (Clusters 1, 3, and 7;  $n = 11$ ,  $n = 26$ , and  $n = 8$  neurons) or licking (Cluster 6;  $n = 7$  neurons) as displayed in Figure 6B, C using a continuous spike density function representation. A comparison between light-responsive and light nonresponse neurons failed to reveal any difference (Fig. 6D,E). Before starting the operant task, all mice had to complete the pretask training to assimilate the cue-reward connection so that the reward was fully predictable. Nevertheless, in 32% (23 of 72 neurons), the response to the predicted reward was larger than the cue response, whereas 44% (32 of 72 neurons) showed the converse (Wilcoxon rank sum test,  $\alpha = 0.05$ ), with only 17% (12 of 72 neurons) having no response at reward consumption during the self-paced operant task (Fig. 6D). Again, the distribution of response amplitudes

measured in optogenetically identified DA neurons (4 of 16 with [reward resp.] > [cue resp.] vs 11 of 16 with [cue resp.] > [reward resp.]; Wilcoxon rank sum test,  $\alpha = 0.05$ ) was similar to responses across all recorded DA neurons (Fig. 6E). Light-nonresponsive VTA neurons of Cluster 5 can thus be considered as putative DA neurons ( $n = 56$ ). The other VTA neurons (i.e., non-Cluster 5 neurons) showed high mean firing rate (>20 Hz), were activated during specific motor output (e.g., licking or locomotion), did not show a clear response at the cue and/or reward, and are hereafter designated as non-DA neurons ( $n = 87$ ). Hereafter, we mainly focused on the DA neurons (Cluster 5,  $n = 72$ ), but we used non-DA neurons to perform various cross-checks.

### Tonic signals reflecting motor activity during cue-reward interval

For DA neurons, spike bursts represent a phasic signal contrasting with sustained spiking that is referred to as the tonic signal. This distinction was formalized as “80/160 ms” criteria (Grace and Bunney, 1984) whereby spikes with an ISI <80 ms marks the beginning of a burst and subsequent spikes are part of the burst until the ISI becomes >160 ms. Here we applied an adaptive threshold method whereby a burst started when the ISI fell below one-third of the mean ISI for a given neuron and ended when two spikes were separated by an ISI longer than two-thirds of the mean ISI. This method yields identical results as the 80/160 rule for spike rates <5 Hz, and has the advantage to accommodate burst with higher frequency. Phasic activity was observed in response to the cue and/or reward (Fig. 3B,D,E) exclusively in the DA neuron cluster (Cluster 5 in Fig. 6A,B). In addition, variation of the tonic activity was observed for some neurons, in particular during the cue-reward interval (Fig. 7A,B). During this period, with the cue lights on, the animal leaves the TZ and runs to the spout to collect the reward, suggesting a possible link to locomotion. To examine this possibility, we aligned the trials to the speed peaks during this period. The mean normalized tonic and pause responses of a fraction of the recorded DA neurons (Fig. 7D) reflect the speed and acceleration modulation (Fig. 7C). Such neural variation of DA neurons could hardly be detected in the auROC profiles aligned to the locomotor activity in Figures 5 and 6 because, there, all speed/acceleration peaks were considered (i.e., inside and outside of the cue-reward interval).

### VTA DA neuronal pause during error trials

We next analyzed the trials where the mouse did not wait 2 s in the TZ (i.e., no cue presented) but still approached the spout and started licking but did not get a reward. We wondered whether we can detect a negative prediction error signal during error trials (Fig. 8A–C). We compared four events compatible with both correct and error trials: the zone entry and exit as well as the LB onset and offset. The zone exit is a proxy for the cue (absent in error trials), whereas the LB offset is a proxy for the reward (also absent in error trials). We considered the neural activity along with the behavioral parameters for all DA neurons ( $n = 72$ ). For a single session, the mean distance to the zone/spout, speed, and licks along with the cue and reward distribution, all aligned to the LB offset, are displayed in Figure 8B. The two speed peaks observed in correct trials are partially merged in error trials since the mouse did not wait the 2 s required to trigger the cue. The firing (spike rate, burst/tonic firing, and pause probability), equally aligned to the LB offset, are shown in Figure 8C for a DA neuron. We observed a negative error signal, associated to the lack of reward, in the form of a significant dip in the spike rate

and tonic firing, but an increase of the pause probability, for error versus correct trials. The absence of burst signal in error trials reflect the absence of the cue and reward. Finally, the firing difference between correct and error trials is unlikely arising from a lower motivation in error trials since the licking profiles are similar in both type of trials (Fig. 8B). To check whether we have a significant modulation of the neural signal at four key time points of the task execution (zone entry/exit and LB onset/offset) between correct and error trials, we resorted to auROC profiles (Fig. 8D; see Materials and Methods) for each firing modality (spike, burst, tonic, and pause). In order to check that the relative change was significant (i.e., that stronger/weaker response to error than to correct trials is significant), we performed a Wilcoxon two-sided rank sum test ( $p < 0.05$ ) and displayed them in a heatmap (Fig. 8D). We also grouped neurons by animal and sorted them by behavioral performance measured as the median inter-reward interval (Fig. 8D) to check that the observed modulations are not associated to the animal's performance. We found spike pause increases in 23 of 56 (41%) of putative DA neurons and 10 of 16 (63%) of optogenetically identified DA neurons for at least one of the four events (Fig. 8D). Discriminating neurons were not necessarily specific to a single event (Fig. 8F). A detailed mean spike rate heatmap of the two main events (leave zone and LB offset) for all DA neurons is provided in Figure 9. In particular, a negative error prediction signal was present at the leave zone event (Fig. 9A), signaling the absence of the cue in the error trials, but in a smaller proportion than for the LB offset (Fig. 9B) signaling the absence of reward. VTA DA neurons can thus exhibit an internally generated error signal aligned to the mouse's actions yielding a distinct activity pattern in the face of very similar motor output.

### Neuronal modulation during goal-directed actions

VTA DA neurons respond to rewarding events, but evidence is accumulating that they are also modulated by motor parameters (Dodson et al., 2016; Da Silva et al., 2018; Engelhard et al., 2019) maybe in conjunction with phasic response to the reward (Puryear et al., 2010). Indeed, we expect this overlapping encoding to be goal-directed. To test this point, we first analyzed the PSTH of each neuron around the locomotion onset, time of maximal speed, and lick onset within and outside the cue-reward window (schematic of their temporal localization is given in Fig. 10A). For example, alignment to the speed peak revealed a change in firing between the cue and reward onset but not when the alignment was made to the time of peak speed outside trials (e.g., Fig. 10B, compare top with bottom panels). We computed significance in a 1 s time window around the motor event for all 159 VTA neurons (Fig. 10C; see Materials and Methods). Up versus down modulation was seen in all VTA neurons; however, only DA neurons showed a clear modulation difference between inside and outside of the cue-reward interval (Fig. 10D; Table 1). For all DA neurons, we also examined the cue and reward responses and found that 53 of 72 showed such phasic responses. Importantly, 39 of 72 (54%) neurons encoded simultaneously elements of motor behavior and cue-reward events (Venn diagram in Fig. 10C). Such overlapping response was more pronounced within the cue-reward window than outside (Fig. 10D). The same analysis was performed excluding all motor events that are closer than 750 ms to the cue or reward leading to similar results (Table 1). These results confirm our previous observation for tonic and pause modulation at the speed peak between cue onset and reward collection (Fig. 7C,D), supporting altogether

the hypothesis that DA neuron modulation encodes motor action when it is goal-directed.

### VTA DA neuron multiplexing in cue-reward intervals

Trying to transmit multiple signals through a single communication channel is commonly referred as multiplexing in telecommunication. The brain, and thus the neurons, can also multiplex (Lankarany et al., 2019). To determine how many variables contribute to the spiking pattern of a single unit, we applied an encoding approach (Fig. 11A) for each neuron using a linear-nonlinear Poisson method, that is, a GLM with an exponential nonlinearity and a Poisson noise (Pillow et al., 2005, 2008; Paninski et al., 2007) (see Materials and Methods). This choice circumvents the possibility of negative spike counts encountered using a Gaussian noise, leading to a purely linear model, when performing spikes encoding/decoding (Triplett and Goodhill, 2019). For a given session, we used seven variables (external events: zone, cue, and reward; as well as motor actions: lick, speed, acceleration, and distance to the goal) and compared the resulting prediction to the recorded spiking through Pearson correlations (Fig. 11B). To compare the encoding during the task execution versus when the animal is not performing the task (Fig. 11C; see Materials and Methods), we computed for each neuron the correlations for all correct trials' time intervals (ranging from the zone entry to the LB offset) versus outside trials' time intervals (excluding error trials where the task is executed but wrongly). Only neurons recorded during session with at least 30 min of uninterrupted recording were used ( $n = 159 \rightarrow 118$ ). In addition, we excluded neurons whose correlation are below chance level to retain only the ones displaying a significant encoding (see Materials and Methods; in brief, we repeated  $1000 \times$  the prediction analysis for temporally shuffled neural signal setting a threshold of 2.807 SDs to the shuffled distribution and found 10 neurons below chance level;  $n = 118 \rightarrow 108$  with 52 non-DA, 43 putative DA, and 13 opto. id. DA; see Fig. 11A). The correlations were overall stronger for non-DA than DA neurons (putative and opto. id. DA) inside and outside of the task. However, for DA neurons ( $n = 43 + 13 = 56$ ), the correlations were significantly higher inside than outside of the task execution. This effect is not context-dependent since the inside versus outside difference for non-DA neurons ( $n = 52$ ) is absent. (Figure 11D,E; Kruskal-Wallis nonparametric unpaired one-way ANOVA with *post hoc* test Tukey-Kramer; KW:  $df = 3$ ,  $\chi^2 = 33.4$ ,  $p = 2.7 \times 10^{-7}$ ; TK:  $\alpha = 0.05$ , e.g., non-DA inside vs outside:  $p(1,2) = 0.50$ , DA inside versus outside:  $p(3,4) = 0.049$ , non-DA inside vs DA inside:  $p(1,3) = 0.016$ , non-DA outside vs DA outside:  $p(2,4) = 0.0003$ ). Together, this indicates that salient, reward-predicting events provide a time window during which goal-directed actions are closely tracked by DA neurons. Finally, we estimated the number of variables significantly contributing to the spiking for each cell. In brief, we performed the encoding analysis for each variables' combination (Fig. 12A, top; see Materials and Methods) by setting the missing variables to their mean value over time for each cell. The number of significantly contributing variables was obtained by selecting the combination with the minimum number of variables necessary to cross 90% of the largest value among the computed maximum correlations (see Materials and Methods; Fig. 12A, bottom). We found that DA neurons typically required three or more contributors to explain their firing pattern based on both external events (i.e., cue and reward, and actions, such as licking or speed) (Fig. 12B, top). On the other hand, non-DA neurons typically encoded two or less variables, such as speed and acceleration (Fig. 12B, top).

The degree of correlation to the variables is overall lower for DA neurons than for non-DA neurons. The reasons find its roots in the very principle of multiplexing. A single channel (the neuron's spiking profile) has to accommodate multiple signals and thus divide its capacity between them. Furthermore, each variable comes with its own noise decreasing the overall encoding capacity. Finally, we most probably only studied a fraction of the parameters, which are multiplexed by the DA neurons in regard of the complexity of the task. Thus, the higher number of encoded variables, noise, and overlooked variables could explain the lower correlation for DA neurons compared with non-DA neurons. This is reflected in Figure 12B where more a neuron encodes variables, the lower the maximum correlation tends to be on average.

To determine the relative importance of the different variables for each group of studied neurons, we computed the fraction of correlation lost by removing one variable at the time (see Materials and Methods). Using a temporal shuffling procedure, we found that the computed loss in correlation of the 108 neurons previously selected is beyond chance level. In addition, we checked that the average loss in correlation is significantly different from zero for all variables and neuronal type independently (for non-DA and DA groups: *t* test vs zero with Bonferroni correction, all  $p < 0.05$ ). Moreover, we identified significant loss in correlation between variables (for non-DA and DA groups: Kruskal-Wallis nonparametric unpaired one-way ANOVA with *post hoc* test Tukey-Kramer; non-DA neurons: KW:  $df = 6$ ,  $\chi^2 = 66.56$ ,  $p = 2.07 \times 10^{-12}$ ; TK:  $\alpha = 0.05$ , e.g., acceleration vs reward:  $p(1,5) = 5.5 \times 10^{-6}$ , speed vs cue:  $p(2,7) = 0.002$ ; DA neurons: KW:  $df = 6$ ,  $\chi^2 = 47.39$ ,  $p = 1.6 \times 10^{-8}$ ; TK:  $\alpha = 0.05$ , e.g., reward vs acceleration:  $p(1,5) = 1.1 \times 10^{-4}$ , reward vs speed:  $p(1,5) = 5.7 \times 10^{-7}$ ). The obtained fractions may not seem large at first (in general,  $\leq 0.25$ ; see, e.g., values for individual neuron examples in Fig. 12C). They remain reasonable since a neuron performing multiplexing has to share its channel between the variables (e.g., four variables would each occupy 25% of the capacity assuming equipartition). As expected, external events (cue and reward) as well as licking were more important for DA neurons, whereas speed and acceleration matter more in non-DA neurons (examples in Fig. 12C and summary in Fig. 12D). The encoding in DA neurons was stronger for the reward than for the cue (Fig. 12D) in accordance to the observation that the neural response was in general stronger to the reward than to the cue (Fig. 6D,E). Furthermore, a DA neuron encoding reward collection did not necessarily also encode licks (see example in Fig. 12C). The less marked differences in loss between variables for DA than non-DA (Fig. 12D) is because DA neurons multiplex more variables (Fig. 12B). In DA neurons, the loss in correlation is smaller for motor-related variables than for reward-related variables (see, e.g., Fig. 12C). From the multiplexing point of view, this originates from the fact that the channel capacity is not equally shared between the variables (the tonic signal varies on a smaller scale than the burst signal). This repartition is biologically meaningful since the most important action to perform a correct trial are to accurately track the cue appearance and the reward collection while the motor output matters to maximize the reward rate. Finally, to control that the largest contributors really matter to the encoding, we compared the distributions of correlations when the encoder was given all seven covariates to when the three strongest covariates were removed (Fig. 12E; two-sample Kolmogorov-Smirnov one-sided test, KS test statistics = 0.381,  $p = 6.1 \times 10^{-9}$ ). Together, the GLM encoding confirmed that multiple variables determine the firing pattern of VTA neurons. More variables contributed in DA

neurons than non-DA neurons. Moreover, the correlations between the variables and the neural signal were higher during than outside of the task execution for DA neurons, but not for non-DA neurons.

## Discussion

Our single-unit recordings in freely moving mice show that, during a self-paced operant task, external events and motor-related activity are encoded simultaneously in the majority of VTA DA neurons using distinct firing modes. The multiplexing was prominent during the task execution in the interval between cue presentation and reward collection.

### Cue and reward coding

A large fraction of neurons responded strongly even to a predicted reward, which is at odds with the classical view of the RPE hypothesis (Schultz et al., 1993). However, in case of predictable rewards, we still observed a response in line with temporal difference algorithms in a standard cue-reward experimental paradigm (Pan et al., 2005). It is unlikely that the incomplete disappearance of a reward response would be because of the introduction of the TZ. Indeed, some DA neurons responded only weakly to the reward during the pretraining phase, where the mice were still learning the cue-reward association. Moreover, this cannot be because of different stages of learning, as all mice had fully associated the cue with the availability of the reward already during the pretraining well before the self-paced operant task was introduced. In line with the two-component phasic DA response model (Schultz, 2016; Watabe-Uchida et al., 2017), the persistent phasic responses to the reward could represent an early saliency signal rather than a late value coding (Nomoto et al., 2010). Indeed, we found that the cue responses contained early saliency (0–250 ms) as well as late value components (250–500 ms), whereas reward responses predominantly occurred during the early 250 ms window (Fig. 6D).

During sessions where several DA neurons were recorded simultaneously, we found some DA neurons showing large responses to the reward along with cue-responsive DA neurons that only showed a very small reward response (Fig. 3D,E). This could be explained by an extension of the classical temporal difference paradigm, namely, distributional reinforcement learning, which has been proposed for single-unit recordings in the VTA of mice performing tasks with probabilistic rewards (Dabney et al., 2020). This model assumes the existence of a diverse set of RPE channels, each of which carries a different value prediction and constitutes a part of a distributional code. Thus, a single reward could elicit a different response in each DA neuron since each of them codes its own value prediction. Similarly, it has been proposed that sensory prediction error is not represented by individual DA neurons but rather encoded in the firing pattern of many DA neurons (Stalnaker et al., 2019). Thus, the diversity in responses observed here would simply reflect the necessity of a multidimensional encoding of prediction error to handle a self-paced complex operant task.

### Correct versus error trials discrimination through spiking pause

In favor of RPE coding, we observed a robust decrease in spiking for error trials, indicating that erroneous actions produced significant decreases in spiking causing a pause. Previous studies have argued that VTA DA neuron pauses represent a negative RPE (Chang et al., 2016; Parker et al., 2016), which in our case

may not be limited to the evaluation of the reward but also encompass actions in general. In our operant task, the animal would build a model of the task structure and continuously improve future outcomes by adjusting their actions (Costa, 2011; Nakahara and Hikosaka, 2012). Another example for DA driving general action refinement may be found in songbirds. When presented with a distorted auditory feedback during song learning, birds produce a similar negative performance signal in VTA DA neurons (Gadagkar et al., 2016).

### Multiplexing by VTA DA neurons

While previous studies using FSCV have argued for task-related variables in addition to RPE coding in DA transients (Howe et al., 2013; Collins et al., 2016; Hamid et al., 2016; Lee et al., 2018), they fell short of identifying individual DA neurons that would encode multiplexed information with high temporal resolution. As mentioned above, the phasic component of DA neuron firing may encode cue and reward, whereas sustained changes in tonic firing in the same cell could encode goal-oriented locomotion (Niv et al., 2007), providing efficient means to transmit complex information through a single channel. In the present study, we found a large fraction of VTA DA neurons whose tonic firing and pauses are concomitant to the speed peak when the animals runs from the TZ to the reward spout (Fig. 7C,D). On the other hand, frequency-division multiplexing has been proposed for DA neuromodulation on theoretical grounds (Niv et al., 2007; Schultz, 2007; Hiroiyuki, 2014; Oster et al., 2015), but experimental evidence remains scarce.

There is also evidence for functionally specialized subpopulations, for example, between midbrain DA neurons of the VTA and SNc. The latter were observed to encode kinematics regardless of the valence of trigger of the movement (Barter et al., 2015). Similarly, monitoring the DA terminals in ventral and dorsal striatum revealed functional specialization (Wang and Tsien, 2011; Howe and Dombeck, 2016). In the VTA, some DA neurons predominantly encoded reward, whereas others mainly encoded movements or both (Puryear et al., 2010; Barter et al., 2015). One study reported burst at the onset and offset of voluntary movement in VTA DA neurons (Wang and Tsien, 2011). Finally, when monitoring the activity of many neurons with calcium imaging in head-fixed mice navigating through a virtual-reality environment, spatially organized subpopulations of DA neurons were identified that may encode information about a subset of behavioral variables in addition to encoding reward (Engelhard et al., 2019).

Using perievent analysis (Da Silva et al., 2018) and a GLM approach (Allen et al., 2017), we found a mixture of external events and goal-directed actions predicting the neural signal of VTA DA neurons (Figs. 10–12) consistent with task-related modulation in DA firing rates described in monkeys while performing a forelimb reaching task (Schultz, 1986). This was in contrast to VTA non-DA neurons whose firing patterns could largely be explained by a small number of motor output (e.g., Fig. 12C, top). The multiplexing of VTA DA during the cue-reward time-interval (or more generally the trial ranging from the zone entry to the LB offset) does not encode precisely all signal, cortical neurons are perfectly fitted to encode with precision motor events (Rickert et al., 2009), but most probably combines them for downstream processing, such as regulating the plasticity of cortical input to the striatum (Wang and Tsien, 2011; Menegas et al., 2015; Howe and Dombeck, 2016). This is in line with optogenetic manipulations where direct stimulation of VTA GABAergic neurons shows disruption of motor output (Van Zessen et al., 2012), whereas phasic firing of VTA DA neurons reinforces action rather than driving motor output (Tsai et al.,

2009; Hamid et al., 2016). In SNc DA neurons, transient increases in firing have been observed before action initiation and optogenetic stimulation promoted movement initiation (Da Silva et al., 2018). Moreover, DA neurons tend to covary with various movement parameters (Dodson et al., 2016; Engelhard et al., 2019).

Multiplexing may arise from the convergence of many motor inputs onto VTA DA neurons, namely, inputs from motor control (M1, M2) and general locomotor regions (e.g., diagonal band of Broca, medial septal nucleus, pedunculopontine nucleus, and laterodorsal tegmentum) (Watabe-Uchida et al., 2012; Fuhrmann et al., 2015; Xiao et al., 2016) and inputs shaping the RPE computation. The latter are believed to be more widely distributed, neurons contributing to the positive RPE signal have been found in the striatum, ventral pallidum, subthalamic nucleus, pedunculopontine nucleus, and lateral hypothalamus (Tian et al., 2016). For the negative error signals, the lateral habenula has emerged as the main input (Matsumoto and Hikosaka, 2007; Tian and Uchida, 2015). Interestingly, specific inputs can differentially affect individual firing modes in DA neurons, thereby precisely addressing a single channel of the multiplexed signal (Floresco et al., 2003). Without surprise, optogenetic manipulations of VTA DA neurons have revealed strong behavioral effects, which may be explained by the many information streams treated in parallel (Lammel et al., 2008, 2011, 2012; Beier et al., 2015; Xiao et al., 2016).

But how might the signal then be demultiplexed? Downstream of the VTA, cell type-specific changes in the NAc can occur during learning (Atallah et al., 2014). Indeed, D1 and D2 receptors expressed on individual medium spiny neurons differentially regulate synaptic plasticity of afferent glutamate transmission (Shen et al., 2008; Yagishita et al., 2014), which has been proposed to underlie reinforcement learning (Schultz, 1998). As D1 and D2 receptors have different affinities, it makes them ideal candidates to decode a multiplexed DA signal (Dreyer et al., 2010; Marcott et al., 2014), and differentiate between the phasic and tonic spiking. In line with this interpretation, does closed-loop stimulation of either D1- or D2-striatal medium spiny neurons during movement execution bidirectionally shift specific parameters of the stimulated movement (Yttri and Dudman, 2016)? Comparable mechanisms could underlie the optimization of task-related parameters in ventral striatum during goal-directed actions. Similar DA-dependent mechanisms have been reported in other downstream areas of the VTA: stimulation of VTA DA projections to PFC showed opposing effects between phasic and tonic stimulation (Ellwood et al., 2017), and functional imaging studies in humans have implicated the posterior medial frontal cortex in learning from errors and showed that it was dopamine and D2 receptor-dependent (Klein et al., 2007).

We provide here experimental evidence for goal-directed multiplexing in individual DA neurons for freely moving mice during an operant task. Cue-reward signals and movement-related firing are processed in parallel thanks to the phasic and tonic modes of DA neurons. Each single unit is modulated in an outcome-dependent manner, thus contributing to the learning of new contingencies in self-paced operant settings. Our results may also have translational implications. Presentation of salient cues to patients with Parkinson's disease can improve symptoms, for example, by overcoming freezing of gait (Ginis et al., 2017; Gilat et al., 2018). If elevated DA levels constitute an underlying mechanism, then our data indicate that adding cues may recruit otherwise silent VTA DA neurons (some of which encode motor actions), which are less affected by neurodegeneration than SNc neurons (Surmeier et al., 2017) and could partially compensate in regions where SNc and VTA projections overlap. Another implication might be that,

in substance abuse, where pharmacological jamming of the error signal by the drug in addition to environmental cues could drive the persistent consumption despite negative consequences and lead to addiction (Keiflin and Janak, 2015; Pascoli et al., 2015).

## References

- Allen WE, Kauvar IV, Chen MZ, Richman EB, Yang SJ, Chan K, Gradinaru V, Deverman BE, Luo L, Deisseroth K (2017) Global representations of goal-directed behavior in distinct cell types of mouse neocortex. *Neuron* 94:891–907.
- Anikeeva P, Andalman AS, Witten I, Warden M, Goshen I, Grosenick L, Gunaydin LA, Frank LM, Deisseroth K (2011) Optetrode: a multichannel readout for optogenetic control in freely moving mice. *Nat Neurosci* 15:163–170.
- Atallah HE, McCool AD, Howe MW, Graybiel AM (2014) Neurons in the ventral striatum exhibit cell-type-specific representations of outcome during learning. *Neuron* 82:1145–1156.
- Barter JW, Li S, Lu D, Bartholomew RA, Rossi MA, Shoemaker CT, Salas-Meza D, Gaidis E, Yin HH (2015) Beyond reward prediction errors: the role of dopamine in movement kinematics. *Front Integr Neurosci* 9:1–22.
- Beier KT, Steinberg EE, DeLoach KE, Xie S, Miyamichi K, Schwarz L, Gao XJ, Kremer EJ, Malenka RC, Luo L (2015) Circuit architecture of VTA dopamine neurons revealed by systematic input-output mapping. *Cell* 162:622–634.
- Blanquat PD, Saint Hok V, Save E, Poucet B, Chaillan FA (2013) Differential role of the dorsal hippocampus, ventro-intermediate hippocampus, and medial prefrontal cortex in updating the value of a spatial goal. *Hippocampus* 23:342–351.
- Brodersen KH, Ong CS, Stephany KE, Buhmann JM (2010) The binormal assumption on precision-recall curves. *Proc Int Conf Pattern Recognit* 4263–4266. Available at <http://ieeexplore.ieee.org/document/5597760/>.
- Chang CY, Esber GR, Marrero-Garcia Y, Yau HJ, Bonci A, Schoenbaum G (2016) Brief optogenetic inhibition of dopamine neurons mimics endogenous negative reward prediction errors. *Nat Neurosci* 19:111–116.
- Coddington LT, Dudman JT (2018) The timing of action determines reward prediction signals in identified midbrain dopamine neurons. *Nat Neurosci* 21:1563–1573.
- Cohen JY, Haesler S, Vogt L, Lowell BB, Uchida N (2012) Neuron-type-specific signals for reward and punishment in the ventral tegmental area. *Nature* 482:85–88.
- Collins AL, Greenfield VY, Bye JK, Linker KE, Wang AS, Wassum KM (2016) Dynamic mesolimbic dopamine signaling during action sequence learning and expectation violation. *Sci Rep* 6:20231.
- Costa RM (2011) A selectionist account of de novo action learning. *Curr Opin Neurobiol* 21:579–586.
- Da Silva JA, Tecuapetla F, Paixão V, Costa RM (2018) Dopamine neuron activity before action initiation gates and invigorates future movements. *Nature* 554:244–248.
- Dabney W, Kurth-Nelson Z, Uchida N, Starkweather CK, Hassabis D, Munos R, Botvinick M (2020) A distributional code for value in dopamine-based reinforcement learning. *Nature* 577:671–675.
- Dodson PD, Dreyer JK, Jennings KA, Syed EC, Wade-Martins R, Cragg SJ, Bolam JP, Magill PJ (2016) Representation of spontaneous movement by dopaminergic neurons is cell-type selective and disrupted in parkinsonism. *Proc Natl Acad Sci USA* 113:E2180–E2188.
- Dreyer JK, Herrik KF, Berg RW, Hounsgaard JD (2010) Influence of phasic and tonic dopamine release on receptor activation. *J Neurosci* 30:14273–14283.
- Ellwood IT, Patel T, Wadia V, Lee AT, Liptak AT, Bender KJ, Sohal VS (2017) Tonic or phasic stimulation of dopaminergic projections to prefrontal cortex causes mice to maintain or deviate from previously learned behavioral strategies. *J Neurosci* 37:1221–1217.
- Engelhard B, Finkelstein J, Cox J, Fleming W, Jang HJ, Ornelas S, Koay SA, Thiberge SY, Daw ND, Tank DW, Witten IB (2019) Specialized coding of sensory, motor and cognitive variables in VTA dopamine neurons. *Nature* 570:509–513.
- Ferguson JE, Boldt C, Redish AD (2009) Creating low-impedance tetrodes by electroplating with additives. *Sens Actuators A Phys* 156:388–393.
- Fiorillo CD, Tobler PN, Schultz W (2003) Discrete coding of reward probability and uncertainty by dopamine neurons. *Science* 299:1898–1902.
- Floresco SB, West AR, Ash B, Moorel H, Grace AA, Moore H, Grace AA (2003) Afferent modulation of dopamine neuron firing differentially regulates tonic and phasic dopamine transmission. *Nat Neurosci* 6:968–973.
- Fuhrmann F, Justus D, Sosulina L, Kaneko H, Beutel T, Friedrichs D, Schoch S, Schwarz MK, Fuhrmann M, Remy S (2015) Locomotion, theta oscillations, and the speed-correlated firing of hippocampal neurons are controlled by a medial septal glutamatergic circuit. *Neuron* 86:1253–1264.
- Gadagkar V, Puzerey PA, Chen R, Baird-Daniel E, Farhang AR, Goldberg JH (2016) Dopamine neurons encode performance error in singing birds. *Science* 354:1278–1282.
- Gilat M, Lúcia Silva de Lima A, Bloem BR, Shine JM, Nonnekes J, Lewis SJ (2018) Freezing of gait: promising avenues for future treatment. *Parkinsonism Relat Disord* 52:7–16.
- Ginis P, Nackaerts E, Nieuwboer A, Heremans E (2017) Cueing for people with Parkinson's disease with freezing of gait: a narrative review of the state-of-the-art and novel perspectives. *Ann Phys Rehabil Med* 61:407–413.
- Grace A, Bunney B (1984) The control of firing pattern in nigral dopamine neurons: burst firing. *J Neurosci* 4:2877–2890.
- Hamid AA, Petibone JR, Mabrouk OS, Hetrick VL, Schmidt R, Vander Weele CM, Kennedy RT, Aragona BJ, Berke JD (2016) Mesolimbic dopamine signals the value of work. *Nat Neurosci* 19:117–126.
- Hanes DP, Thompson KG, Schall JD (1995) Relationship of presaccadic activity in frontal eye field and supplementary eye field to saccade initiation in macaque: Poisson spike train analysis. *Exp Brain Res* 103:85–96.
- Hill DN, Mehta SB, Kleinfeld D (2011) Quality metrics to accompany spike sorting of extracellular signals. *J Neurosci* 31:8699–8705.
- Hiroiyuki N (2014) Multiplexing signals in reinforcement learning with internal models and dopamine. *Curr Opin Neurobiol* 25:123–129.
- Hok V, Chah E, Save E, Poucet B (2013) Prefrontal cortex focally modulates hippocampal place cell firing patterns. *J Neurosci* 33:3443–3451.
- Hollerman JR, Schultz W (1998) Dopamine neurons report an error in the temporal prediction of reward during learning. *Nat Neurosci* 1:304–309.
- Howe MW, Dombeck DA (2016) Rapid signalling in distinct dopaminergic axons during locomotion and reward. *Nature* 535:505–510.
- Howe MW, Tierney PL, Sandberg SG, Phillips PE, Graybiel AM (2013) Prolonged dopamine signalling in striatum signals proximity and value of distant rewards. *Nature* 500:575–579.
- Keiflin R, Janak PH (2015) Dopamine prediction errors in reward learning and addiction: from theory to neural circuitry. *Neuron* 88:247–263.
- Klein TA, Neumann J, Reuter M, Hennig J, von Cramon DY, Ullsperger M (2007) Genetically determined differences in learning from errors. *Science* 318:1642–1645.
- Lammel S, Hetzel A, Häckel O, Jones I, Liss B, Roeper J (2008) Unique properties of mesoprefrontal neurons within a dual mesocorticolimbic dopamine system. *Neuron* 57:760–773.
- Lammel S, Ion DI, Roeper J, Malenka RC (2011) Projection-specific modulation of dopamine neuron synapses by aversive and rewarding stimuli. *Neuron* 70:855–862.
- Lammel S, Lim BK, Ran C, Huang KW, Betley MJ, Tye KM, Deisseroth K, Malenka RC (2012) Input-specific control of reward and aversion in the ventral tegmental area. *Nature* 491:212–217.
- Lankarany M, Al-Basha D, Ratté S, Prescott SA (2019) Differentially synchronized spiking enables multiplexed neural coding. *Proc Natl Acad Sci USA* 116:10097–10102.
- Lee B, Gentry RN, Bissonette GB, Herman RJ, Mallon JJ, Bryden DW, Calu DJ, Schoenbaum G, Coutureau E, Marchand AR, Khamassi M, Roesch MR (2018) Manipulating the revision of reward value during the inter-trial interval increases sign tracking and dopamine release. *PLoS Biol* 16:e2004015.
- London TD, Licholai JA, Szczot I, Ali MA, Le Blanc KH, Fobbs WC, Kravitz AV (2018) Coordinated ramping of dorsal striatal pathways preceding food approach and consumption. *J Neurosci* 38:3547–3558.
- Marcott PF, Mamaligas AA, Ford CP (2014) Phasic dopamine release drives rapid activation of striatal D2 receptors. *Neuron* 84:164–176.
- Matsumoto M, Hikosaka O (2007) Lateral habenula as a source of negative reward signals in dopamine neurons. *Nature* 447:1111–1115.
- Menegas W, Bergan JF, Ogawa SK, Isogai Y, Venkataraju KU, Osten P, Uchida N, Watabe-Uchida M (2015) Dopamine neurons projecting to the posterior striatum form an anatomically distinct subclass. *Elife* 4:1–30.

- Mohebi A, Pettibone JR, Hamid AA, Wong JM, Vinson LT, Patriarchi T, Tian L, Kennedy RT, Berke JD (2019) Dissociable dopamine dynamics for learning and motivation. *Nature* 570:65–70.
- Nakahara H, Hikosaka O (2012) Learning to represent reward structure: a key to adapting to complex environments. *Neurosci Res* 74:1–7.
- Niv Y, Daw ND, Joel D, Dayan P (2007) Tonic dopamine: opportunity costs and the control of response vigor. *Psychopharmacology (Berl)* 191:507–520.
- Nomoto K, Schultz W, Watanabe T, Sakagami M (2010) Temporally extended dopamine responses to perceptually demanding reward-predictive stimuli. *J Neurosci* 30:10692–10702.
- Oster A, Faure P, Gutkin BS (2015) Mechanisms for multiple activity modes of VTA dopamine neurons. *Front Comput Neurosci* 9:1–17.
- Pan WX, Schmidt R, Wickens JR, Hyland BI (2005) Dopamine cells respond to predicted events during classical conditioning: evidence for eligibility traces in the reward-learning network. *J Neurosci* 25:6235–6242.
- Paninski L, Pillow J, Lewi J (2007) Statistical models for neural encoding, decoding, and optimal stimulus design. *Prog Brain Res* 165:493–507.
- Parker NF, Cameron CM, Taliaferro JP, Lee J, Choi JY, Davidson TJ, Daw ND, Witten IB (2016) Reward and choice encoding in terminals of mid-brain dopamine neurons depends on striatal target. *Nat Neurosci* 19:845–854.
- Pascoli V, Terrier J, Hiver A, Lüscher C (2015) Sufficiency of mesolimbic dopamine neuron stimulation for the progression to addiction. *Neuron* 88:1054–1066.
- Paxinos G, Franklin KB (2001) The mouse brain in stereotaxic coordinates, Ed 2. San Diego: Academic.
- Phillips PE, Stuber GD, Heien MV, Wightman RM, Carelli RM (2003) Subsecond dopamine release promotes cocaine seeking. *Nature* 422:614–618.
- Pillow JW, Paninski L, Uzzell VJ, Simoncelli EP, Chichilnisky EJ (2005) Prediction and decoding of retinal ganglion cell responses with a probabilistic spiking model. *J Neurosci* 25:11003–11013.
- Pillow JW, Shlens J, Paninski L, Sher A, Litke AM, Chichilnisky EJ, Simoncelli EP (2008) Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature* 454:995–999.
- Puryear CB, Kim MJ, Mizumori SJY (2010) Conjunctive encoding of movement and reward by ventral tegmental area neurons in the freely navigating rodent. *Behav Neurosci* 124:234–247.
- Quiroga RQ, Nadasdy Z, Ben-Shaul Y (2004) Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering. *Neural Comput* 16:1661–1687.
- Rickert J, Riehle A, Aertens A, Rotter S, Nawrot MP (2009) Dynamic encoding of movement direction in motor cortical neurons. *J Neurosci* 29:13870–13882.
- Roitman MF, Stuber GD, Phillips PE, Wightman RM, Carelli RM (2004) Dopamine operates as a subsecond modulator of food seeking. *J Neurosci* 24:1265–1271.
- Sadacca BF, Jones JL, Schoenbaum G (2016) Midbrain dopamine neurons compute inferred and cached value prediction errors in a common framework. *Elife* 5:1–13.
- Saunders BT, Richard JM, Margolis EB, Janak PH (2018) Dopamine neurons create Pavlovian conditioned stimuli with circuit-defined motivational properties. *Nat Neurosci* 21:1072–1083.
- Schmitzer-Torbert N, Jackson J, Henze D, Harris K, Redish AD (2005) Quantitative measures of cluster quality for use in extracellular recordings. *Neuroscience* 131:1–11.
- Schultz W (1986) Responses of midbrain dopamine neurons to behavioral trigger stimuli in the monkey. *J Neurophysiol* 56:1439–1461.
- Schultz W (1998) Predictive reward signal of dopamine neurons. *J Neurophysiol* 80:1–27.
- Schultz W (2007) Multiple dopamine functions at different time courses. *Annu Rev Neurosci* 30:259–288.
- Schultz W (2016) Dopamine reward prediction-error signalling: a two-component response. *Nat Rev Neurosci* 17:183–195.
- Schultz W, Apicella P, Ljungberg T (1993) Responses of monkey dopamine neurons to reward and conditioned stimuli during successive steps of learning a delayed response task. *J Neurosci* 13:900–913.
- Schultz W, Dayan P, Montague PR (1997) A neural substrate of prediction and reward. *Science* 275:1593–1599.
- Shen W, Flajolet M, Greengard P, Surmeier DJ (2008) Dichotomous dopaminergic control of striatal synaptic plasticity. *Science* 321:848–851.
- Soares S, Atallah BV, Paton JJ (2016) Midbrain dopamine neurons control judgment of time. *Science* 354:1273–1277.
- Stalnaker TA, Howard JD, Takahashi YK, Gershman SJ, Kahnt T, Schoenbaum G (2019) Dopamine neuron ensembles signal the content of sensory prediction errors. *Elife* 8:723908.
- Starkweather CK, Babayan BM, Uchida N, Gershman SJ (2017) Dopamine reward prediction errors reflect hidden-state inference across time. *Nat Neurosci* 20:581–589.
- Surmeier DJ, Obeso JA, Halliday GM (2017) Selective neuronal vulnerability in Parkinson disease. *Nat Rev Neurosci* 18:101–113.
- Syed EC, Grima LL, Magill PJ, Bogacz R, Brown P, Walton ME (2016) Action initiation shapes mesolimbic dopamine encoding of future rewards. *Nat Neurosci* 19:34–36.
- Tian J, Huang R, Cohen JY, Osakada F, Kobak D, Machens CK, Callaway EM, Uchida N, Watabe-Uchida M (2016) Distributed and mixed information in monosynaptic inputs to dopamine neurons. *Neuron* 91:1374–1389.
- Tian J, Uchida N (2015) Habenula lesions reveal that multiple mechanisms underlie dopamine prediction errors. *Neuron* 87:1304–1316.
- Tobler PN, Fiorillo CD, Schultz W (2005) Adaptive coding of reward value by dopamine neurons. *Science* 307:1642–1645.
- Triplet MA, Goodhill GJ (2019) Probabilistic encoding models for multivariate neural data. *Front Neural Circuits* 13:1.
- Tsai HC, Zhang F, Adamantidis A, Stuber GD, Bonci A, de Lecea L, Deisseroth K (2009) Phasic firing in dopaminergic neurons is sufficient for behavioral conditioning. *Science* 324:1080–1084.
- Van Zessen R, Phillips JL, Budygin EA, Stuber GD (2012) Activation of VTA GABA neurons disrupts reward consumption. *Neuron* 73:1184–1194.
- Wallisch P, Lusignan M, Benayoun M, Baker T, Dickey A, Hatsopoulos N (2014) MATLAB for neuroscientists. Amsterdam: Elsevier.
- Walton ME, Bouret S (2019) What is the relationship between dopamine and effort? *Trends Neurosci* 42:79–91.
- Wang DV, Tsien JZ (2011) Conjunctive processing of locomotor signals by the ventral tegmental area neuronal population. *PLoS One* 6:e16528.
- Watabe-Uchida M, Zhu L, Ogawa SK, Vamanrao A, Uchida N (2012) Whole-brain mapping of direct inputs to midbrain dopamine neurons. *Neuron* 74:858–873.
- Watabe-Uchida M, Eshel N, Uchida N (2017) Neural circuitry of reward prediction error. *Annu Rev Neurosci* 40:373–394.
- Xiao C, Cho JR, Zhou C, Treweek JB, Chan K, McKinney SL, Yang B, Gradinaru V (2016) Cholinergic mesopontine signals govern locomotion and reward through dissociable midbrain pathways. *Neuron* 90:333–347.
- Yagishita S, Hayashi-Takagi A, Ellis-Davies GC, Urakubo H, Ishii S, Kasai H (2014) A critical time window for dopamine actions on the structural plasticity of dendritic spines. *Science* 345:1616–1620.
- Yttri EA, Dudman JT (2016) Opponent and bidirectional control of movement velocity in the basal ganglia. *Nature* 533:402–406.